

AUTOMORPHIC FUNCTIONS

BY
LESTER R. FORD, Ph.D.
*Professor of Mathematics
Illinois Institute of Technology*

This volume was written during the tenure of a
National Research Fellowship

EA libris G.T.

SECOND EDITION

CHELSEA PUBLISHING COMPANY
231 WEST 29TH STREET
NEW YORK 1, N. Y.

1951

COPYRIGHT 1929, BY CHELSEA PUBLISHING COMPANY

REPRINTED WITH CORRECTIONS, 1951

PRINTED IN THE UNITED STATES OF AMERICA

PREFACE TO THE FIRST EDITION

In 1915, the author published in the series of *Edinburgh Mathematical Tracts* a brief introduction to "Automorphic Functions." This booklet has been long out of print. Except for this little volume, no book on the subject has ever appeared in English. This is regrettable, in view of the importance of the subject to those whose interests lie in the field of Functions of a Complex Variable and of its numerous contacts with other domains of mathematical thought.

It has been the author's aim in the earlier chapters to lay the foundations of the theory with all possible rigor and simplicity. The introduction and use of the isometric circle (the name was suggested by Professor Whittaker) have given the theory of linear groups a simplicity it has not had hitherto. The fundamental region which results was given by J. I. Hutchinson, in 1907, and independently by G. Humbert, in 1919. But the interesting properties of the circle itself and its utility in the derivation of the theory seem to have escaped the attention of workers in the field. It may be stated here, that much of the material involving the isometric circle embodies researches of the author which have not appeared elsewhere in published form.

In the later chapters, the author's task has been largely a matter of selection of material and method of treatment. Here, also, the use of the isometric circle has often led to a simplification of the proofs. The material included has been, perhaps, a matter of the author's personal taste. The classical elliptic modular functions deserve a place as the best-known examples of non-elementary automorphic functions. The theory of conformal mapping has been presented at some length—from the modern function-theory point of view—as a preparation for the theories of uniformization which follow. The chapter on Conformal Mapping can be read independently of the rest of the book. In the final chapter, the connection between automorphic functions and differential equations is brought out.

PREFACE

This treatment, which is necessarily brief, leads up to the triangle functions.

The connection between groups and non-Euclidean geometry has not been treated, as it now seems of less importance, in view of the way the foundations have been laid. A chapter on Abelian Integrals, treated in the light of uniformization, would have been of interest; however, the subject has not been worked out to any extent. This chapter should probably not be written until some geometer sets up a product for the prime function for the Fuchsian group of the first kind. Finally, the book might have been improved by a more extensive use of the theory of normal families of functions.

It is now almost fifty years since Poincaré created the general theory of automorphic functions, in a brilliant series of papers in the early volumes of *Acta Mathematica*. Since that time, the subject has had a steady growth. The material in the present volume will be found to spring very largely from researches of the past twenty years, either in content or in method of treatment. The theory of uniformization rests on the papers of Koebe and Poincaré, published in 1907. The first rigorous proof of the possibility of mapping one plane simply connected region upon another was presented by Osgood, in 1901, but the treatment of this and similar problems by function-theory methods came a dozen years later. Area theorems and other aspects of mapping are of still later date and are active subjects of investigation today. The foundations of the theory of groups, as previously stated, are based on the author's studies during the past few years.

The author wishes to thank Prof. E. T. Whittaker for various kindnesses while at the Mathematical Laboratory of the University of Edinburgh. He is indebted similarly to Profs. Otto Hölder and Paul Koebe, of the University of Leipzig. He received much profit from the lectures of Professor Hölder on "Elliptic Modular Functions" and from those of Professor Koebe on "Uniformization." So much of the material in the latter half of the book was gleaned from the published papers of Professor Koebe that specific references have often not been given and attention is called here to that fact. The author acknowledges a debt of long standing to Prof. W. F. Osgood, whose inspiring teaching first aroused an interest in the subject dealt with in this volume and whose "Funktionentheorie" has been a mine of usefulness.

PREFACE

The preceding gentlemen are in no wise responsible for such defects as the book possesses. Their encouragement and their suggestions have been valuable; but they are busy men, and the author has not presumed to ask them to read the manuscript. The only assistance in the task of writing the book was from Mr. Jack Kronsbein, a student of the University of Leipzig, who, as a labor of friendship, typed most of the manuscript and checked many of the formulæ.

The author's chief debts are to the National Research Council, whose grant of a fellowship made the writing of the book possible, and to the Rice Institute, which gave him leave of absence.

L. R. F.

HOUSTON, TEXAS.
June, 1929.

PREFACE TO THE SECOND EDITION

This volume has been out of print for several years and I have had many letters asking how it could be secured. I am much pleased therefore that the Chelsea Publishing Company has wished to reissue the book in its valuable series of reprints.

There has been no essential revision of the text. Numerous errors have been brought to my attention by readers, particularly in the formulas, and these have been corrected.

L. R. F.

CHARLOTTESVILLE, VIRGINIA
August 1951.

CONTENTS

	PAGE
PREFACE.	V

CHAPTER I

LINEAR TRANSFORMATIONS

1. The Linear Transformation	1
2. Symbolic Notation	4
3. The Fixed Points of the Transformation.	6
4. The Linear Transformation and the Circle.	8
5. Inversion in a Circle	10
6. The Multiplier, K	15
7. The Hyperbolic Transformation, $K = A$	18
8. The Elliptic Transformation, $K = e^{i\theta}$	19
9. The Loxodromic Transformation, $K = Ae^{i\theta}$	20
10. The Parabolic Transformation.	21
11. The Isometric Circle	23
12. The Unit Circle	30

CHAPTER II

GROUPS OF LINEAR TRANSFORMATIONS

13. Definition of a Group. Examples	33
14. Properly Discontinuous Groups	35
15. Transforming a Group	36
16. The Fundamental Region.	37
17. The Isometric Circles of a Group.	39
18. The Limit Points of a Group.	41
19. Definition of the Region R	44
20. The Regions Congruent to R	44
21. The Boundary of R	47
22. Example. A Finite Group	49
23. Generating Transformations.	50
24. Cyclic Groups.	51
25. The Formation of Groups by the Method of Combination.	56
26. Ordinary Cycles	59
27. Parabolic Cycles.	62
28. Function Groups.	64

CHAPTER III

FUCHSIAN GROUPS

29. The Transformations.	67
30. The Limit Points.	67

CONTENTS

31. The Region R and the Region R_0	PAGE 69
32. Generating Transformations	71
33. The Cycles	72
34. Fuchsian Groups of the First and Second Kinds	73
35. Fixed Points at Infinity. Extension of the Method	75
36. Examples	78
37. The Modular Group	79
38. Some Subgroups of the Modular Group	81

CHAPTER IV

AUTOMORPHIC FUNCTIONS

39. The Concept of the Automorphic Function	83
40. Simple Automorphic Functions	86
41. Behavior at Vertices and Parabolic Points	88
42. The Poles and Zeros	91
43. Algebraic Relations	94
44. Differential Equations	98

CHAPTER V

THE POINCARÉ THETA SERIES

45. The Theta Series	102
46. The Convergence of the Series	104
47. The Convergence for the Fuchsian Group of the Second Kind	106
48. Some Properties of the Theta Functions	108
49. Zeros and Poles of the Theta Functions	112
50. Series and Products Connected with the Group	115

CHAPTER VI

THE ELEMENTARY GROUPS

I. The Finite Groups

51. Inversion in a Sphere	117
52. Stereographic Projection	119
53. Rotations of the Sphere	120
54. Groups of the Regular Solids	123
55. A Study of the Cube	124
56. The General Regular Solid	127
57. Determination of All the Finite Groups	129
58. The Extended Groups	136

II. The Groups with One Limit Point

59. The Simply and Doubly Periodic Groups	139
60. Groups Allied to the Periodic Groups	140
61. The Automorphic Functions	144

III. The Groups with Two Limit Points

62. Determination of the Groups	146
---	-----

CONTENTS

CHAPTER VII

THE ELLIPTIC MODULAR FUNCTIONS

63. Certain Results from the Theory of Elliptic Functions	PAGE 148
64. Change of the Primitive Periods	150
65. The Function $J(\tau)$	151
66. Behavior of $J(\tau)$ at the Parabolic Points	153
67. Further Properties of $J(\tau)$	155
68. The Function $\lambda(\tau)$	157
69. The Relation between $\lambda(\tau)$ and $J(\tau)$	159
70. Further Properties of $\lambda(\tau)$	160

CHAPTER VIII

CONFORMAL MAPPING

71. Conformal Mapping	164
72. Schwarz's Lemma	165
73. Area Theorems	167
74. The Mapping of a Circle on a Plane Finite Region	169
75. The Deformation Theorem for the Circle	171
76. A General Deformation Theorem	175
77. An Application of Poisson's Integral	177
78. The Mapping of a Plane Simply Connected Region on a Circle. The Iterative Process	179
79. The Convergence of the Process	183
80. The Behavior of the Mapping Function on the Boundary	187
81. Regions Bounded by Jordan Curves	198
82. Analytic Arcs and the Continuation of the Mapping Function across the Boundary	201
83. Circular Arc Boundaries	202
84. The Mapping of Combined Regions	203
85. The Mapping of Limit Regions	205
86. The Mapping of Simply Connected Finite-sheeted Regions	213
87. Conformal Mapping and Groups of Linear Transformations	216

CHAPTER IX

UNIFORMIZATION. ELEMENTARY AND FUCHSIAN FUNCTIONS

88. The Concept of Uniformization	220
89. The Connectivity of Regions	221
90. Algebraic Functions of Genus Zero. Uniformization by Means of Rational Functions	229
91. Algebraic Functions of Genus Greater than Zero. Uniformiza- tion by Means of Automorphic Functions	233
92. The Genus of the Fundamental Region of a Group	238
93. The Cases $p = 1$ and $p > 1$	239
94. More General Fuchsian Uniformizing Functions	241
95. The Case $p = 0$	245

CONTENTS

96. Whittaker's Groups.	PAGE 247
97. The Transcendental Functions.	249

CHAPTER X

UNIFORMIZATION. GROUPS OF SCHOTTKY TYPE

98. Regions of Planar Character.	256
99. Some Accessory Functions.	258
100. The Mapping of a Multiply Connected Region of Planar Character on a Slit Region.	262
101. Application to the Uniformization of Algebraic Functions.	266
102. A Convergence Theorem.	267
103. The Sequence of Mapping Functions.	271
104. The Linearity of T_n	273
105. An Extension.	278
106. The Mapping of a Multiply Connected Region of Planar Character on a Region Bounded by Complete Circles.	279

CHAPTER XI

DIFFERENTIAL EQUATIONS

107. Connection with Groups of Linear Transformations.	284
108. The Inverse of the Quotient of Two Solutions.	287
109. Regular Singular Points of Differential Equations.	293
110. The Quotient of Two Solutions at a Regular Singular Point.	296
111. Equations with Rational Coefficients.	299
112. The Equation with Two Singular Points.	303
113. The Hypergeometric Equation.	303
114. The Riemann-Schwarz Triangle Functions.	305
115. Equations with Algebraic Coefficients.	308

A BIBLIOGRAPHY OF AUTOMORPHIC FUNCTIONS.	311
--	-----

AUTHOR INDEX.	325
-----------------------	-----

SUBJECT INDEX.	327
------------------------	-----

AUTOMORPHIC FUNCTIONS

CHAPTER I

LINEAR TRANSFORMATIONS

1. The Linear Transformation.—Let z and z' be two complex numbers connected by some functional relation, $z' = f(z)$. Let the values of z be represented in the customary manner on an Argand diagram, or z -plane, and the values of z' be represented on a second Argand diagram, or z' -plane. To each point z of the first plane for which the function is defined there correspond one or more values of z' by virtue of the functional relation. To points, curves, and areas of the z -plane there correspond, usually, points, curves, and areas in the z' -plane. We shall speak of the configurations in the z -plane as being *transformed* by the functional relation into the corresponding configurations in the z' -plane.

We shall find it convenient to represent z' and z on the same Argand diagram, rather than on different ones. Then the functional relation transforms configurations in the z -plane into other configurations in the z -plane. In what follows but one plane will be used unless the contrary is stated.

The whole theory of automorphic functions depends upon a particular type of transformation, defined as follows:

DEFINITION.—The transformation

$$z' = \frac{az + b}{cz + d}, \quad (1)$$

where a, b, c, d are constants and $ad - bc \neq 0$, is called a linear transformation.¹

The present chapter will be devoted to a study of this fundamental transformation.

¹ This is more properly called a "linear fractional transformation"; but we shall use the briefer designation. It is also called a "homographic transformation."

If $ad - bc = 0$, the equation reduces to $z' = \text{constant}$; but this case is without interest.

The quantity $ad - bc$ is called the *determinant* of the transformation. It will be convenient to have always

$$ad - bc = 1, \quad (2)$$

the determinant in the general case becoming 1 if the numerator and denominator of the fraction in the second member be divided by $\pm \sqrt{ad - bc}$.

The second member of (1) is an analytic function of z . The linear transformation has, therefore, the property of conformality; that is, when a figure is transformed, angles are preserved both in magnitude and in sign.

We note that for each value of z , equation (1) gives one and only one value of z' . There is no exception to this statement if we introduce the point at infinity. Thus if $c \neq 0$, $z = -d/c$ is transformed into $z' = \infty$, and $z = \infty$ into $z' = a/c$; if $c = 0$, $z = \infty$ is transformed into $z' = \infty$.

Let equation (1) be solved for z :

$$z = \frac{-dz' + b}{cz' - a}. \quad (3)$$

This transformation which, applied after the transformation (1) has been made, carries each configuration back into its original position is called the *inverse* of the transformation (1). We note that (3) is a linear transformation; hence,

THEOREM 1.—*The inverse of a linear transformation is a linear transformation.*

We note that (3) is formed from (1) by interchanging a and d with a change of sign. When formed in this way the determinant is the same as in (1).

We see from (3) that to each value of z' there corresponds one and only one value of z . We have, then, the following result:

THEOREM 2.—*The z -plane is transformed into itself in a one-to-one manner by a linear transformation.*

Moreover, the linear transformation is the most general analytic transformation which has the property stated in Theorem 2. We shall prove first the following theorem:

THEOREM 3.—*If, except for a finite number of points, the plane is mapped in a one-to-one and directly conformal manner upon a plane region, the mapping function is linear.*

Let $z' = f(z)$ be such a mapping function; and let $q_1, q_2, \dots, q_n (= \infty)$ be the excepted points. Owing to the conformality, $f(z)$ is analytic except at the isolated points $q_1, \dots, q_n$. Now q_1

is not an essential singularity, else the function takes on certain values an infinite number of times in the neighborhood of the point, which is contrary to hypothesis. Hence, $f(z)$ either remains finite in the neighborhood of q_1 , and hence is analytic there if properly defined, or has a pole. So $f(z)$ is a rational function of z .

A rational function which is not a constant takes on every value m times, where m is the number of its poles. Since $f(z)$ takes on no value twice, it has a single pole of the first order. If the pole is at a finite point q_k we may write

$$z' = \frac{A_1}{z - q_k} + A_0 = \frac{A_0z + A_1 - A_0q_k}{z - q_k}, \quad A_1 \neq 0. \quad (4)$$

If the pole is at infinity, we have

$$z' = A_1z + A_0, \quad A_1 \neq 0. \quad (4')$$

In either case the function is linear.

COROLLARY 1.—*The most general one-to-one and directly conformal (where conformality has a meaning) transformation of the plane into itself is a linear transformation.*

We have not defined conformality when one of the points involved is the point at infinity. Excepting the point $z = \infty$ and the point of the z -plane which is carried into $z' = \infty$, the transformation is to be conformal. Theorem 3 then applies.

COROLLARY 2.—*The most general one-to-one and directly conformal transformation of the finite plane into itself is the linear transformation $z' = A_1z + A_0$.*

We shall now consider the successive performance of linear transformations. After subjecting the z -plane to the transformation (1) let a second linear transformation

$$z'' = \frac{\alpha z' + \beta}{\gamma z' + \delta} \quad (5)$$

be made. Expressing z'' as a function of z , we have

$$z'' = \frac{\alpha \frac{az + b}{cz + d} + \beta}{\gamma \frac{az + b}{cz + d} + \delta} = \frac{(\alpha a + \beta c)z + \alpha b + \beta d}{(\gamma a + \delta c)z + \gamma b + \delta d}. \quad (6)$$

Making the transformation (1) and then making (5) is equivalent to making the single transformation (6). Now, (6) is a linear transformation; its determinant, in the form in which the fraction is written, is $(ad - bc)(\alpha\delta - \beta\gamma)$. It is worth noting that if the

determinants of (1) and (5) are each unity, that of (6) is also unity without further change.

If z'' be subjected to a linear transformation, we conclude on combining the new transformation with (6) that the succession of three linear transformations is equivalent to a single linear transformation, and so on. We have then the following result:

THEOREM 4.—*The successive performance of a finite number of linear transformations is equivalent to a single linear transformation.*

A further well-known property of the linear transformation is expressed in the following theorem:

THEOREM 5.—*The linear transformation leaves invariant the cross-ratio of four points.*

Let z_1, z_2, z_3, z_4 be four distinct points and let z_1', z_2', z_3', z_4' be the points into which they are transformed by (1). We shall suppose all the points are finite. We have

$$z_1' - z_2' = \frac{az_1 + b}{cz_1 + d} - \frac{az_2 + b}{cz_2 + d} = \frac{(ad - bc)(z_1 - z_2)}{(cz_1 + d)(cz_2 + d)} \quad (7)$$

$$= \frac{z_1 - z_2}{(cz_1 + d)(cz_2 + d)},$$

whence

$$\frac{(z_1' - z_2')(z_3' - z_4')}{(z_1' - z_3')(z_2' - z_4')} = \frac{(z_1 - z_2)(z_3 - z_4)}{(z_1 - z_3)(z_2 - z_4)} \quad (8)$$

If one of the points is at infinity, we make the necessary change in (8) by a limiting process. Thus, if $z_2 = \infty$ and $z_1' = \infty$, (8) becomes

$$\frac{z_3' - z_4'}{z_2' - z_4'} = -\frac{z_3 - z_4}{z_1 - z_3}$$

2. Symbolic Notation.—For brevity in writing and for convenience in combination, we shall represent the second member of a transformation such as (1) by the functional notation, using capital letters for the function; thus,

$$T(z) = \frac{az + b}{cz + d}$$

so that (1) becomes

$$z' = T(z).$$

We shall speak of this as the transformation T , the argument z being omitted unless ambiguity might arise without it. If two transformations are the same, $T_1(z) = T(z)$, this is indicated by the equation $T_1 = T$.

Let S be the transformation (5), so

$$z'' = S(z').$$

Then, (6) is

$$z'' = S[T(z)] = ST(z).$$

Thus, the succession of two transformations is written as a product, the enclosing brackets of the functional notation being omitted. It should be noted that ST is the single linear transformation resulting from making first the transformation T and then the transformation S , the order of performance being from right to left. TS is, in general, different from ST .

It is easily seen from the meaning of the symbols that the associative law of multiplication holds,

$$U(ST) = (US)T,$$

and there is no ambiguity in writing simply UST . In a product, any sequence of factors may be combined into a single linear transformation.

The transformations equivalent to performing T twice, thrice, etc., are represented by T^2, T^3 , etc. Thus, $T^2(z)$ means $T[T(z)]$. The inverse of T is written T^{-1} ; hence, from (3)

$$T^{-1}(z) = \frac{-dz + b}{cz - a}.$$

The result of applying the inverse n times is represented by T^{-n} . If we represent the identical transformation, $z' = z$, by 1 so that $1(z)$ means z , we observe that positive and negative integral powers of T together with unity combine in accordance with the law of the addition of exponents in multiplication.

The inverse of a sequence of transformations can now be written down. To find the inverse of ST we make on the plane transformed by ST the transformation S^{-1} followed by T^{-1} ; we have

$$(T^{-1}S^{-1})(ST) = T^{-1}(S^{-1}S)T = T^{-1}T = 1.$$

Thus, $T^{-1}S^{-1}$ is the transformation which, applied after ST , carries each point back to its original position; so $T^{-1}S^{-1}$ is the inverse of ST . In a similar manner we have for any number of transformations

$$(ST \dots UV)^{-1} = V^{-1}U^{-1} \dots T^{-1}S^{-1}.$$

The rule for the transposition of factors by division is easily found. Let $UST = V$; then,

$$(UST)T^{-1} = VT^{-1}, \text{ or } US = VT^{-1},$$

and

$$U^{-1}(UST) = U^{-1}V, \text{ or } ST = U^{-1}V.$$

Then, in an equation connecting two products, the first (last) factor of one member can be transferred to the beginning (end) of the other member by changing the sign of its exponent. Thus, symbolic division is permissible, provided the proper order of performing the operations is followed. For example, the inverse W of $ST \dots UV$ is a transformation such that

$$WST \dots UV = 1.$$

By repeated division on the right we get the result given above.

3. The Fixed Points of the Transformation.—The points which are unchanged by the transformation (1) are found by setting $z' = z$ in (1) and solving the resulting equation.

$$z = \frac{az + b}{cz + d}, \text{ or } cz^2 + (d - a)z - b = 0. \quad (9)$$

Suppose, first, that $c \neq 0$. Then, (9) has the two roots

$$\xi_1, \xi_2 = \frac{a - d \pm \sqrt{M}}{2c} \quad (10)$$

where

$$M = (d - a)^2 + 4bc = (a + d)^2 - 4. \quad (11)$$

The second expression for M is derived on the assumption that $ad - bc = 1$. We see from (1) that ∞ is not transformed into itself, so there are at most two fixed points. If $M = 0$, that is if $a + d = \pm 2$, there is but one fixed point,

$$\xi = \frac{a - d}{2c}. \quad (12)$$

If $c = 0$, we must have $a \neq 0$, $d \neq 0$ since, otherwise, the determinant would be zero. We see from (1) that ∞ is then a fixed point. Solving (9) we get a finite fixed point provided $a \neq d$. The fixed points are

$$\xi_1 = \frac{b}{d - a}, \quad \xi_2 = \infty. \quad (13)$$

If $c = 0$ and $a = d$, (1) takes the form

$$z' = z + b',$$

a translation with the single fixed point, $\xi = \infty$. In the case $c = 0$, we see from (11) that we have, as before, two fixed points if $M \neq 0$, and one fixed point if $M = 0$.

There cannot be more than two fixed points unless (9) is identically zero; that is, $c = 0$, $d = a$, and $b = 0$. Equation (1) then takes the form $z' = z$. Hence,

THEOREM 6.—The only linear transformation with more than two fixed points is the identical transformation $z' = z$.

By means of this theorem we are able to prove the following important proposition:

THEOREM 7.—There is one and only one linear transformation which transforms three distinct points, z_1, z_2, z_3 , into three distinct points, z_1', z_2', z_3' .

We shall prove first that there is not more than one such transformation. Let T be one transformation carrying z_1, z_2, z_3 into z_1', z_2', z_3' , and let S be any other such transformation. Consider the transformation $T^{-1}S$. We have $S(z_1) = z_1'$ and $T^{-1}(z_1') = z_1$; so

$$T^{-1}S(z_1) = T^{-1}(z_1') = z_1.$$

Hence, z_1 , and similarly z_2 and z_3 , are fixed points of the transformation $T^{-1}S$. It follows from Theorem 6 that

$$T^{-1}S = 1;$$

whence applying the transformation T to both members,

$$S = T.$$

There is, thus, not more than one transformation of the kind required.

We shall prove that there is always one such transformation by actually setting it up. If none of the six values is infinite, consider the transformation defined by

$$\frac{(z' - z_1')(z_2' - z_3')}{(z' - z_2')(z_1' - z_3')} = \frac{(z - z_1)(z_2 - z_3)}{(z - z_2)(z_1 - z_3)}, \quad (14)$$

an equation which expresses the equality of the cross-ratios $(z'z_1', z_2'z_3')$ and (zz_1, z_2z_3) . This is of the form (1) when solved for z' in terms of z . It obviously transforms z_1, z_2, z_3 into z_1', z_2', z_3' ; for both members of (14) are equal to zero when $z = z_1, z' = z_1'$; they are both infinite when $z = z_2, z' = z_2'$; and they are both 1 when $z = z_3, z' = z_3'$.

If one of the given points is at infinity, we have but to replace the member of (14) in which that point occurs by its limiting value when the required variable becomes infinite. If $z_1 = \infty, z_2 = \infty$, or $z_3 = \infty$, we replace the second member of (14) by

$$-\frac{z_2 - z_3}{z - z_2}, \quad -\frac{z - z_1}{z_1 - z_3}, \quad \text{or} \quad \frac{z - z_1}{z - z_2},$$

respectively; and a similar change is necessary in the first member for an infinite value of z_1', z_2' , or z_3' . In any case, there is one

transformation with the desired property, and the theorem is established.

Equation (14) is a convenient form for use in actually setting up the transformation carrying three given points into three given points. Theorem 7 will be of great utility in our subsequent work; to prove that two transformations are identical, we shall have merely to show that they transform three points in the same way.

4. The Linear Transformation and the Circle.—Since we are operating on the complex number z it will be convenient to have the equations of curves expressed directly in terms of z . If x is the real part of z and iy is its imaginary part, and if we represent by $\bar{z}$ the conjugate imaginary of z , we have

$$z = x + iy, \quad \bar{z} = x - iy. \quad (15)$$

From these we have

$$x = \frac{1}{2}(z + \bar{z}), \quad y = \frac{i}{2}(\bar{z} - z), \quad z\bar{z} = x^2 + y^2. \quad (16)$$

From the first two equations of (16) we can readily express the equation of any curve in terms of z and $\bar{z}$.

We shall now get the general equation of the circle and of the straight line. The equation

$$A(x^2 + y^2) + b_1x + b_2y + C = 0,$$

where the constants are real, is the general equation of the circle (possibly of imaginary or zero radius) if $A \neq 0$, and is the general equation of the straight line if $A = 0$ and b_1 and b_2 are not both zero. Substituting from (16) we have

$$Az\bar{z} + \frac{1}{2}(b_1 - ib_2)z + \frac{1}{2}(b_1 + ib_2)\bar{z} + C = 0.$$

Putting $B = \frac{1}{2}(b_1 - ib_2)$, this takes the form

$$Az\bar{z} + Bz + \bar{B}\bar{z} + C = 0 \quad (17)$$

where A and C are real. Equation (17) is the general equation of the circle if $A \neq 0$ and of the straight line if $A = 0$, $B \neq 0$.

The center and radius of the circle are easily found. Writing (17) in the form

$$\left(z + \frac{\bar{B}}{A}\right)\left(\bar{z} + \frac{B}{A}\right) = \frac{B\bar{B} - AC}{A^2},$$

we see that the first member is the square of the distance of z from $-\bar{B}/A$. Hence, (17) is a circle with center $-\bar{B}/A$ and

radius $\sqrt{\frac{B\bar{B} - AC}{A^2}}$. Since we shall be interested only in real circles, we shall require that $B\bar{B} > AC$.

Now let us see what the circle or straight line (17) becomes when the linear transformation (1) is applied. Substituting from (3)

$$z = \frac{-dz' + b}{cz' - a}, \quad \bar{z} = \frac{-\bar{d}\bar{z}' + \bar{b}}{\bar{c}\bar{z}' - \bar{a}},$$

we have

$$A \frac{(-dz' + b)(-\bar{d}\bar{z}' + \bar{b})}{(cz' - a)(\bar{c}\bar{z}' - \bar{a})} + B \frac{-dz' + b}{cz' - a} + \bar{B} \frac{-\bar{d}\bar{z}' + \bar{b}}{\bar{c}\bar{z}' - \bar{a}} + C = 0. \quad (18)$$

Clearing of fractions and collecting terms, we have

$$\begin{aligned} [A\bar{d}d - B\bar{c}d - \bar{B}c\bar{d} + Cc\bar{c}]z'\bar{z}' \\ + [-A\bar{b}d + B\bar{a}d + B\bar{b}c - C\bar{a}c]z' \\ + [-A\bar{b}d + B\bar{a}d + B\bar{b}c - C\bar{a}c]\bar{z}' \\ + A\bar{b}b - B\bar{a}b - B\bar{a}b + C\bar{a}a = 0. \end{aligned} \quad (19)$$

In this equation the coefficient of $z'\bar{z}'$ is real, for $d\bar{d}$ and $c\bar{c}$ are real, being each the product of a number by its conjugate; and $B\bar{c}d + \bar{B}c\bar{d}$ is real, being the sum of a number and its conjugate. Similarly the constant term is real. Also the coefficient of $\bar{z}'$ is the conjugate of the coefficient of z' . Therefore (19) is of the form (17). We have then the following result:

THEOREM 8.—The linear transformation carries a circle or straight line into a circle or straight line.

It will often be convenient to consider the straight line as a circle of infinite radius, in which case we say briefly that a circle is carried into a circle.

It is easy to see when the transform will be a straight line. The straight line is characterized by the fact that it passes through the point ∞ . Hence, if the point which is carried to ∞ lies on the original circle or straight line, the transform will be a straight line; otherwise the transform will be a circle. This is easily shown analytically. For the coefficient of $z'\bar{z}'$ will vanish if (dividing by $c\bar{c}$)

$$A \left(-\frac{d}{c}\right) \left(-\frac{\bar{d}}{\bar{c}}\right) + B \left(-\frac{d}{c}\right) + \bar{B} \left(-\frac{\bar{d}}{\bar{c}}\right) + C = 0;$$

that is, if $-d/c$ lies on the original circle or straight line.

Since three points determine a circle, we can set up a transformation which carries three distinct points of the first circle into three distinct points of the second circle. Having chosen

the first three points, the transformed points can be selected in an infinite variety of ways (∞^3 ways, in fact); and each different selection gives a different transformation. Hence,

THEOREM 9.—*There exist infinitely many linear transformations which transform a given circle into a second given circle.*

In particular, we may choose the second circle to be the same as the first. Hence, there are ∞^3 linear transformations which transform a given circle into itself.

5. Inversion in a Circle.—There is an intimate relation, as we shall now show, between the linear transformation of the complex variable and the geometrical transformation known as "inversion in a circle."

Consider a circle Q with center at K and radius r . Let P be any point of the plane and construct the half line KP , beginning at K and passing through P .

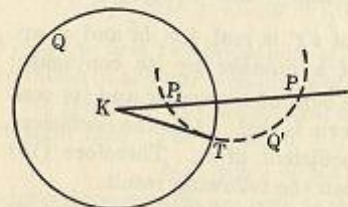


FIG. 1.

Let P_1 be a point on the half line KP such that $KP_1 \cdot KP = r^2$; then P_1 is called the "inverse of P with respect to the circle Q ." The relation is a reciprocal one; P is the inverse of P_1 . We speak of P and P_1 as points inverse with respect to Q .

Inverse points have the property that any circle passing through P and P_1 , the inverse of P with respect to Q , is orthogonal to Q . For, let Q' be any circle through P and P_1 and draw KT tangent to Q' , T being the point of tangency. We have

$$KT^2 = KP_1 \cdot KP = r^2,$$

whence, T lies on Q . The radii to T are perpendicular and the circles are orthogonal.

We shall now get an analytic expression for the transformation. Let P, P_1 , and K be the points z, z_1 , and k in the Argand diagram. The equations of the transformations are

$$|(z_1 - k)(z - k)| = r^2, \quad \arg(z_1 - k) = \arg(z - k).$$

The first equation expresses the condition $KP_1 \cdot KP = r^2$; the second expresses the collinearity of K, P , and P_1 . Since $\arg(z - k) = -\arg(\bar{z} - \bar{k})$, the two equations are satisfied if, and only if,

$$(z_1 - k)(\bar{z} - \bar{k}) = r^2. \quad (20)$$

This is the equation of inversion in terms of complex variables.

If Q is the circle,

$$Az\bar{z} + Bz + \bar{B}\bar{z} + C = 0, \quad (17)$$

equation (20) becomes, on substituting the center and radius previously found,

$$\left(z_1 + \frac{\bar{B}}{A}\right)\left(\bar{z} + \frac{B}{A}\right) = \frac{B\bar{B} - AC}{A^2},$$

which on simplification is

$$Az_1\bar{z} + Bz_1 + \bar{B}\bar{z} + C = 0, \quad (21)$$

We thus get the relation between z and its inverse z_1 from the equation of Q by substituting z_1 for z and leaving $\bar{z}$ unchanged. Solving (21) we have the explicit form of the transformation,

$$z_1 = \frac{-\bar{B}\bar{z} - C}{A\bar{z} + B}. \quad (22)$$

When $A = 0$ so that (17) is a straight line, we shall still use formulæ (21) and (22) for the inversion. It is not difficult to show geometrically that when A approaches zero, P and P_1 attain positions such that Q is the perpendicular bisector of the segment PP_1 . Inversion then becomes a reflection in the line Q . To show this analytically, let z_2 be a point on Q . Then, $|z_2 - z|$ is the distance from z_2 to z . The distance from z_2 to z_1 , using the equation of the transformation and the equation of Q (with $A = 0$) which z_2 satisfies, is

$$|z_2 - z_1| = \left| \frac{-\bar{B}z_2 - C}{B} + \frac{\bar{B}z + C}{B} \right| = \left| \frac{\bar{B}}{B} (\bar{z} - \bar{z}_2) \right| = |z_2 - z|.$$

Thus all points of the line Q are equidistant from P and P_1 .

We shall now prove the following proposition:

THEOREM 10.—*The linear transformation carries two points which are inverse with respect to a circle into two points which are inverse with respect to the transformed circle.*

Let z and z_1 be inverse with respect to the circle (17); then (21) is satisfied. Make the transformation (1) and let z', z'_1 be the transformed points. We have

$$z_1 = \frac{-dz'_1 + b}{cz'_1 - a}, \quad \bar{z} = \frac{-\bar{d}\bar{z}' + \bar{b}}{\bar{c}\bar{z}' - \bar{a}};$$

whence, substituting in (21),

$$A \frac{(-dz'_1 + b)(-\bar{d}\bar{z}' + \bar{b})}{(cz'_1 - a)(\bar{c}\bar{z}' - \bar{a})} + B \frac{-dz'_1 + b}{cz'_1 - a} + B \frac{-\bar{d}\bar{z}' + \bar{b}}{\bar{c}\bar{z}' - \bar{a}} + C = 0.$$

This equation is the same as (18) except that z' is replaced by z_1' ; hence, on simplifying we shall get (19) with $\bar{z}'$ unchanged and z' replaced by z_1' . But this is the condition that z' and z_1' be inverse points with respect to the transformed circle (19).

Let us return to our study of the inversion. We see from (22) that to each z there corresponds one and only one z_1 . Likewise to each z_1 corresponds one and only one z ; hence, the transformation is one-to-one. The fixed points of the inversion are seen, from the geometrical construction, to be the points of Q itself. This appears analytically if we set $z_1 = z$ in (21).

The inversion (22) can be written as the succession of the two transformations

$$z_2 = \bar{z}, \quad z_1 = \frac{-\bar{B}z_2 - C}{Az_2 + B}.$$

The first of these is a reflection in the real axis; the second is a linear transformation. The first preserves the magnitudes of angles but reverses their signs; the second makes no alteration. Hence, the inversion is inversely conformal. This follows also from the fact that the second member of (22) is an analytic function of $\bar{z}$.

The reflection in the real axis obviously carries circles into circles, and the succeeding linear transformation does likewise. Again the reflection carries a circle and two points which are inverse with respect to it into a circle and two inverse points; and this relation, by Theorem 10, is preserved by the linear transformation.

We summarize the results in the following theorem:

THEOREM 11.—*Inversion in a circle is a one-to-one inversely conformal transformation which carries circles into circles, and carries two points inverse with respect to a circle into two points inverse with respect to the transformed circle.*

Since the inversion is a one-to-one transformation of the plane which preserves the magnitude of the angle but changes its sign, the result of performing two inversions, or any even number of inversions, is a one-to-one transformation which preserves both the magnitude and the sign of the angle. According to Theorem 3, Corollary 1, such a transformation is a linear transformation. Hence, we have the theorem:

THEOREM 12.—*The successive performance of an even number of inversions is equivalent to a linear transformation.*

We shall prove that, conversely, any linear transformation is so constituted. First, let us examine some of the simpler linear transformations and find equivalent pairs of inversions.

(a) *The translation, $z' = z + b$.*

By this transformation each point of the plane is translated parallel to the line Ob (Fig. 2) a distance equal to the length of Ob . Let L_1 and L_2 be two lines perpendicular to the line Ob and at a distance apart equal to half the length of Ob . A reflection in L_1 followed by a reflection in L_2 , the lines being designated as in the figure, is equivalent to the given translation. It is sufficient to note that three points are transformed in the proper manner (Theorem 7). We observe at once that

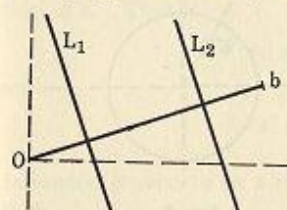


FIG. 2.

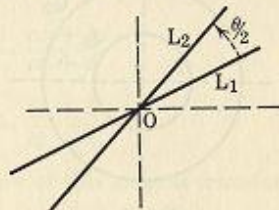


FIG. 3.

the points of L_1 , which are unchanged by the first reflection, are translated in the desired manner by the second reflection.

(b) *The rotation, $z' = e^{i\theta}z$.*

Each point is rotated about the origin through an angle θ . A reflection in L_1 , followed by a reflection in L_2 , arranged as in Fig. 3, clearly rotates the points of L_1 as required; hence, the two reflections are equivalent to the desired rotation.

(c) *The stretching from the origin, $z' = Az$, $A > 0$.*

Each point is transformed into a point with the same argument, but with the modulus multiplied by A . There is a stretching from the origin, or, if $A < 1$, a contraction toward the origin. This is equivalent to an inversion in a circle Q_1 with center at the origin and radius r_1 (Fig. 4) followed by an inversion in a circle Q_2 with center at the origin and radius $r_2 = r_1\sqrt{A}$. For, if z_1, z' are the successive transforms of z , we have

$$z_1\bar{z} = r_1^2, \quad z'\bar{z}_1 = r_1^2A;$$

whence,

$$z' = \frac{r_1^2A}{\bar{z}_1} = r_1^2A \cdot \frac{z}{r_1^2} = Az.$$

(d) *The transformation $z' = -1/z$.*

This can be written

$$z_1 = -\bar{z}, \quad z' = \frac{1}{\bar{z}_1}.$$

It is, thus, a reflection in the imaginary axis, $z + \bar{z} = 0$, followed by an inversion in the unit circle, $z\bar{z} = 1$ (Fig. 5).

Consider now the general transformation (1). If $c \neq 0$, we can write this

$$z' - \frac{a}{c} = \frac{bc - ad}{c(cz + d)} = -\frac{1}{c^2 \left(z + \frac{d}{c} \right)}, \quad (22')$$

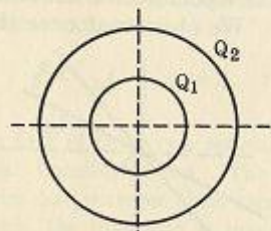


FIG. 4.

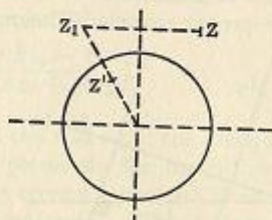


FIG. 5

supposing $ad - bc = 1$. This can be written as the following sequence of transformations:

$$z_1 = z + \frac{d}{c}, \quad z_2 = cz_1, \quad z_3 = -\frac{1}{z_2}, \quad z' = z_4 + \frac{a}{c}.$$

The first, third, and fourth of these transformations have each been shown to be equivalent to a pair of inversions. The second can be broken into two, putting $c^2 = Ae^{i\theta}$, $A > 0$,

$$z_2 = e^{i\theta} z_1, \quad z_3 = Az_2,$$

each of which is equivalent to a pair of inversions.

If $c = 0$, the transformation has the form

$$z' = az + \beta.$$

Putting $\alpha = Ae^{i\theta}$, this is equivalent to the following sequence:

$$z_1 = e^{i\theta} z, \quad z_2 = Az_1, \quad z' = z_2 + \beta.$$

We have proved the converse of Theorem 12.

THEOREM 13.—Any linear transformation is equivalent to the successive performance of an even number of inversions in circles.

A linear transformation can be expressed as the sequence of inversions in an infinite number of ways. Later, we shall show

that any linear transformation is equivalent to four suitably chosen inversions, and that a transformation of the class subsequently called "non-loxodromic" can be expressed as a succession of two inversions.

Let us consider now the most general one-to-one inversely conformal transformation of the plane,

$$z' = \bar{V}(z),$$

where, of course, $\bar{V}$ is not an analytic function of z . If we first make a reflection in the real axis, $z_1 = \bar{z}$, and then apply the preceding, we have a one-to-one directly conformal transformation, and hence, by Theorem 3, Corollary 1, a linear transformation of z . That is,

$$\bar{V}(z_1) = \bar{V}(\bar{z}) = \frac{az + b}{cz + d},$$

and

$$z' = \bar{V}(z) = \frac{a\bar{z} + b}{c\bar{z} + d}. \quad (23)$$

Inversion is seen to be a special case of this general transformation.

Equation (23) is a reflection followed by a linear transformation. So we can state the following general result:

THEOREM 14.—The most general one-to-one conformal transformation of the plane into itself is equivalent to a succession of inversions in circles. The transformation is directly or inversely conformal according as the number of inversions is even or odd.

6. The Multiplier, K.—We have already noted a separation of linear transformations into two classes. Putting aside the identical transformation $z' = z$, a transformation has either two fixed points or one fixed point. The number of fixed points and the behaviour of the transformation with reference to the fixed points furnish a useful basis of classification of linear transformations. We treat, first, the larger class with two fixed points.

Suppose, first, that in the transformation (1) $c \neq 0$. The finite points ξ_1, ξ_2 [Equation (10)] and the point ∞ are carried into ξ_1, ξ_2 , and a/c , respectively. Hence, from (14), the transformation may be written

$$\frac{(z' - \xi_1) \left(\frac{a}{c} - \xi_2 \right)}{(z' - \xi_2) \left(\frac{a}{c} - \xi_1 \right)} = \frac{z - \xi_1}{z - \xi_2},$$

or

$$\frac{z' - \xi_1}{z' - \xi_2} = K \frac{z - \xi_1}{z - \xi_2}, \quad (24)$$

where

$$K = \frac{a - c\xi_1}{a - c\xi_2}. \quad (25)$$

K is called the "multiplier" of the transformation; its value, as we shall see, determines the character of the transformation.

To get an expression for K in terms of the coefficients of the transformation, we form the following symmetric function of ξ_1 and ξ_2 :

$$K + \frac{1}{K} = \frac{a - c\xi_1}{a - c\xi_2} + \frac{a - c\xi_2}{a - c\xi_1} = \frac{2a^2 - 2ac(\xi_1 + \xi_2) + c^2(\xi_1^2 + \xi_2^2)}{a^2 + ac(\xi_1 + \xi_2) + c^2\xi_1\xi_2}.$$

Since ξ_1 and ξ_2 are the roots of

$$cz^2 + (d - a)z - b = 0, \quad (9)$$

we have

$$\xi_1 + \xi_2 = -\frac{a-d}{c}, \quad \xi_1\xi_2 = -\frac{b}{c}, \quad \xi_1^2 + \xi_2^2 = \frac{(a-d)^2 + 2bc}{c^2}.$$

Making these substitutions and simplifying, we have

$$K + \frac{1}{K} = \frac{(a+d)^2 - 2ad + 2bc}{ad - bc}.$$

or, if $ad - bc = 1$,

$$K + \frac{1}{K} = (a+d)^2 - 2. \quad (26)$$

We observe from (26) that the value of K depends solely upon the value of $a + d$. If we replace K by $1/K$, equation (26) is unaltered; hence, the two roots of (26) are reciprocals. The particular root to be used in (24) depends upon which fixed point is called ξ_1 and which ξ_2 .

Another simple equation satisfied by K is got from (26) by transposing the 2 and extracting the square root:

$$\sqrt{K} + \frac{1}{\sqrt{K}} = a + d. \quad (27)$$

We now make the change of variables

$$Z = G(z) = \frac{z - \xi_1}{z - \xi_2}, \quad Z' = G(z') = \frac{z' - \xi_1}{z' - \xi_2}, \quad (28)$$

transformations which carry ξ_1 and ξ_2 to 0 and ∞ , respectively.

Then, (24) takes the form

$$Z' = KZ. \quad (29)$$

Call this transformation K so that $K(Z) = KZ$, and we have for the original transformation

$$z' = G^{-1}(Z') = G^{-1}K(Z) = G^{-1}KG(z).$$

Writing for the original transformation $z' = T(z)$, we have

$$T = G^{-1}KG, \text{ whence } K = GTG^{-1}. \quad (30)$$

Let F be any configuration (point, circle, region, or what not) and let F be carried into F' by K . Operating with T on $G^{-1}(F)$, we have

$$TG^{-1}(F) = G^{-1}KGG^{-1}(F) = G^{-1}K(F) = G^{-1}(F');$$

that is, T carries $G^{-1}(F)$ into $G^{-1}(F')$. We shall use this fact in the following manner: We shall investigate the simple transformation (29) and, having found how configurations are transformed, we shall carry the results back to the case of z and z' by applying G^{-1} .

The transformation with $c = 0$,

$$z' = \frac{az + b}{d}, \quad ad = 1$$

can be put in the form (29) and treated similarly. One fixed point is ∞ , the other is $\xi_1 = \frac{b}{d-a}$. We find easily that

$$z' - \xi_1 = K(z - \xi_1) \quad (31)$$

where

$$K = \frac{a}{d} \quad (32)$$

Putting

$$Z = G(z) = z - \xi_1, \quad Z' = G(z') = z' - \xi_1,$$

transformations which carry ξ_1 and ∞ to 0 and ∞ , respectively, we have (29) as before. We have also

$$K + \frac{1}{K} = \frac{a}{d} + \frac{d}{a} = \frac{a^2 + d^2}{ad} = \frac{(a+d)^2 - 2ad}{ad} = (a+d)^2 - 2;$$

hence, for this case also K satisfies equations (26) and (27).

An advantage in writing a transformation in terms of K lies in the ease with which powers of the transformation can be written down. If the transformation (1) be repeated n times, the equivalent single transformation becomes rapidly complicated if expressed in terms of a, b, c, d . But, if we use (24) or (31), we

have, obviously, as the result of n applications of the transformation

$$\frac{z' - \xi_1}{z' - \xi_2} = K^n \frac{z - \xi_1}{z - \xi_2}, \text{ or } z' - \xi_1 = K^n (z - \xi_1). \quad (33)$$

Thus, we merely replace the multiplier K by K^n . Similarly, for the inverse we use the multiplier K^{-1} ; and for n applications of the inverse we use K^{-n} .

Writing K in terms of its modulus A (>0) and its amplitude θ , so

$$K = Ae^{i\theta},$$

we distinguish the three classes of transformations treated in the following sections.

7. The Hyperbolic Transformation, $K = A$.—We assume that $A \neq 1$, since, otherwise, we have the identical transformation. The transformation $Z' = AZ$ is the stretching from the

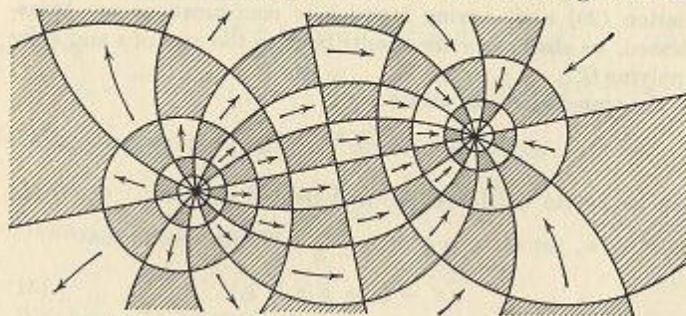


FIG. 6.

origin studied in Sec. 5(c). We observe at once the following facts concerning it: (1) a straight line through the origin (that is, a circle through the fixed points 0 and ∞) is transformed into itself, each half line issuing from the origin being transformed into itself; (2) the half plane on one side of a line through 0 is transformed into itself; (3) any circle with center at the origin (and hence orthogonal to the family of fixed lines) is transformed into some other circle with center at the origin; (4) the points 0 and ∞ are inverse points with respect to any circle with center at the origin.

We now make the transformation G^{-1} which carries 0 and ∞ to ξ_1 and ξ_2 . We have, then, the following facts concerning the hyperbolic transformation: (1) any circle through the fixed

points is transformed into itself, each of the two arcs into which the circle is separated by the fixed points being transformed into itself; (2) the interior of a circle through the fixed points is transformed into itself; (3) any circle orthogonal to the circles through the fixed points is carried into some other such circle; (4) the fixed points are inverse with respect to each circle of (3).

Figure 6 shows the two families of circles just mentioned. The way in which regions are transformed is indicated in the figure, each shaded region being transformed into the next in the direction of the arrow.

From (26), we get the condition that the transformation be hyperbolic in terms of $a + d$. The quantity $K + \frac{1}{K}$ has for real positive values of K the minimum value 2 when $K = 1$. Since $K \neq 1$, we have $K + \frac{1}{K} > 2$; whence, from (26), $(a + d)^2 > 4$. Hence, in order that the transformation be hyperbolic it is necessary that $a + d$ be real and $|a + d| > 2$. That this condition is sufficient will appear presently.

8. The Elliptic Transformation, $K = e^{i\theta}$.—Here, $\theta \neq 2n\pi$. The transformation $Z' = e^{i\theta}Z$ is the rotation about the origin

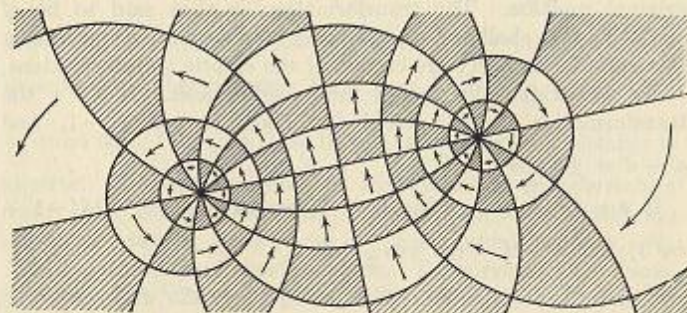


FIG. 7.

discussed in Sec. 5(b). The straight lines and circles of the preceding section have their rôles interchanged. The circle with center at the origin is transformed into itself, the interior of the circle being transformed into itself. The points 0 and ∞ are inverse with respect to each fixed circle. A line drawn through the origin is transformed into a line through the origin which makes an angle θ with the first line.

Applying the transformation G^{-1} , which carries 0 and ∞ to ξ_1 and ξ_2 , respectively, we have the following facts: (1) an arc of a circle joining the fixed points is transformed into an arc of a circle joining the fixed points and making an angle θ with the first arc; (2) each circle orthogonal to the circles through the fixed points is transformed into itself; (3) the interior of each circle of (2) is transformed into itself; (4) the fixed points are inverse points with respect to each circle of (2).

The character of the transformation, with $\theta = \frac{1}{2}\pi$ is shown in Fig. 7. The shaded regions are transformed into shaded regions as indicated by the arrows.

For the elliptic transformation, (26) gives

$$(a + d)^2 = 2 + e^{i\theta} + e^{-i\theta} = 2 + 2 \cos \theta. \quad (34)$$

The second member is positive or zero and less than 4. Hence, $a + d$ is real and $|a + d| < 2$. From (27) we have

$$a + d = \pm \left(e^{i\frac{\theta}{2}} + e^{-i\frac{\theta}{2}} \right) = \pm 2 \cos \frac{\theta}{2}. \quad (35)$$

If θ is commensurable with π , there will exist an integer n such that $n\theta = 2m\pi$; and $K^n = e^{2\pi mi} = 1$. The result of applying the transformation n times is that each point is returned to its original position. The transformation is then said to be of period n . We shall find that the only transformations possessing this periodic property are certain of the elliptic transformations.

We illustrate with two or three useful cases. If $\theta = \pi$ the transformation is of period 2. Then $K = e^{\pi i} = -1$, and $a + d = \pm 2 \cos \frac{\pi}{2} = 0$.

If $\theta = 2\pi/3$, the period is 3. Then $K = e^{2\pi i/3} = \frac{1}{2}(-1 + i\sqrt{3})$, and $a + d = \pm 2 \cos \frac{\pi}{3} = \pm 1$.

If $\theta = \pi/2$, the period is 4. Then $K = e^{\pi i/2} = i$, and $a + d = \pm 2 \cos \frac{\pi}{4} = \pm \sqrt{2}$.

9. The Loxodromic Transformation, $K = Ae^{i\theta}$.—Here A is positive and unequal to 1 and $\theta \neq 2n\pi$. The transformation $Z' = Ae^{i\theta}Z$ can be written as the succession of the transformations

$$Z' = e^{i\theta}Z_1, \quad Z_1 = AZ,$$

of which one is hyperbolic, the other elliptic. There is a stretching from the origin followed by a rotation about the origin. Each

circle with center at the origin is carried into another circle with center at the origin; and each half line through the origin is carried into a half line making an angle θ with the first.

For the original transformation, there is a combination of the motions shown in Figs. 6 and 7. Each circular arc joining the fixed points is carried into another such arc making an angle θ with the first. Each circle orthogonal to the circles through the fixed points is carried into another such orthogonal circle.

The loxodromic transformation has, in general, no fixed circles. There is an exception when $\theta = \pi$. Then, any circular arc joining the fixed points is carried into another arc joining the fixed points and making an angle π with the first, the two arcs thus forming a circle. Any circle through the fixed points is then carried into itself. There is, however, this difference from the preceding cases: *the interior of a fixed circle is transformed into its exterior.*

For the loxodromic transformation, (26) gives

$$\begin{aligned} (a + d)^2 &= 2 + Ae^{i\theta} + \frac{1}{A}e^{-i\theta} \\ &= 2 + \left(A + \frac{1}{A}\right) \cos \theta + i \left(A - \frac{1}{A}\right) \sin \theta. \end{aligned} \quad (36)$$

In general, the second member is not real. If, however, $\theta = \pi$ the second member becomes $2 - \left(A + \frac{1}{A}\right)$, which is real. But

we found that $A + \frac{1}{A} > 2$; so in this case the second member is negative. Without exception in the loxodromic transformation $a + d$ is a complex (non-real) number. In the loxodromic transformation with fixed circles $a + d$ is a pure imaginary.

10. The Parabolic Transformation.—There remains the transformation with one fixed point, which is called a "parabolic transformation." The condition that there is a single fixed point (Sec. 3) is that $a + d = \pm 2$. Then the multiplier, if defined by (26), has the value $K = 1$.

If $c \neq 0$, the fixed point is $\xi = (a - d)/2c$. The transformation carries ∞ , ξ , $-d/c$ into a/c , ξ , ∞ , respectively; hence, by the use of the formula (14), it can be written

$$\frac{z' - \frac{a}{c}}{z' - \xi} = \frac{\xi + \frac{d}{c}}{z - \xi}.$$

Subtracting 1 from each member,

$$\frac{\xi - \frac{a}{c}}{z' - \xi} = \frac{\xi + \frac{d}{c}}{z - \xi} - 1.$$

Now,

$$\begin{aligned}\xi - \frac{a}{c} &= \frac{a-d}{2c} - \frac{a}{c} = -\frac{a+d}{2c} = \mp \frac{1}{c}, \\ \xi + \frac{d}{c} &= \frac{a-d}{2c} + \frac{d}{c} = \frac{a+d}{2c} = \pm \frac{1}{c}.\end{aligned}$$

Hence, the transformation can be written in the form

$$\frac{1}{z' - \xi} = \frac{1}{z - \xi} \pm c. \quad (37)$$

In (37), we have $+c$ if $a+d=2$, and $-c$ if $a+d=-2$

Making the change of variable,

$$Z = G(z) = \frac{1}{z - \xi}, \quad Z' = G(z') = \frac{1}{z' - \xi}, \quad (38)$$

a transformation which carries ξ to ∞ , we have

$$Z' = Z \pm c. \quad (39)$$

If $c=0$, so that ∞ is the single fixed point, we have already found, in Sec. 3, that the transformation is of the form (39) without further change. We have, in fact, $a=d=\pm 1$, and

$$z' = z \pm b. \quad (39')$$

The transformation (39) is the translation discussed in Sec. 5(a). The plane is translated parallel to the line joining the origin to the point $\pm c$. Any line parallel to this line is transformed into itself. The half plane on one side of a fixed line is carried into itself. Any other straight line is carried into a parallel line.

On applying G^{-1} , ∞ is carried to ξ . Parallel straight lines, intersecting at ∞ only, are carried into circles intersecting only at ξ , and, hence, tangent at ξ . Hence, in the parabolic transformation: (1) any circle through the fixed point is transformed into a tangent circle through the fixed point; (2) there is a one parameter family of tangent circles each of which is transformed into itself; (3) the interior of each fixed circle is transformed into itself.

The manner in which the plane is transformed is shown in Fig. 8. Each shaded region is carried in the direction of the arrow.

It is clear from the Figs. 6 to 8 and from the reasoning on which they are based that if a linear transformation is hyperbolic, elliptic, or parabolic there passes through each point of the plane, other than a fixed point, a unique fixed circle. In particular, there is in each case a single fixed circle through ∞ ; that is, there is one fixed straight line. This line is easily constructed; for it passes through the point $-d/c$ which is carried to ∞ and the point a/c into which ∞ is carried.

At this point we shall combine certain of the results of the latter sections into a theorem for reference. We exclude the identical transformation $z' = z$.

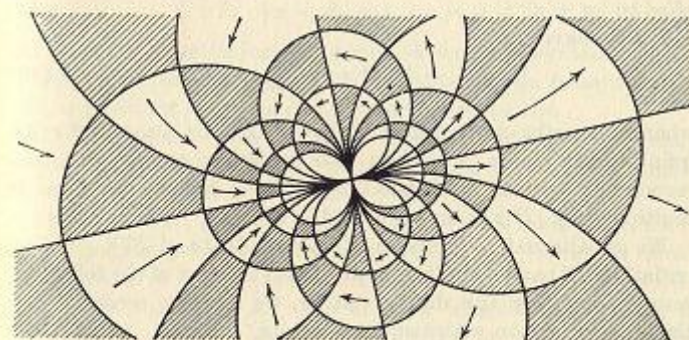


FIG. 8.

THEOREM 15.—The transformation $z' = (az + b)/(cz + d)$, where $ad - bc = 1$, is of the type stated if, and only if, the following conditions on $a + d$ hold:

Hyperbolic, if $a + d$ is real and $|a + d| > 2$.

Elliptic, if $a + d$ is real and $|a + d| < 2$.

Parabolic, if $a + d = \pm 2$.

Loxodromic, if $a + d$ is complex.

We have proved that these conditions are necessary. That they are sufficient follows, by elementary reasoning, from the fact that they are mutually exclusive. Thus, if $a + d$ is real and $|a + d| > 2$, the transformation can be neither elliptic, parabolic, nor loxodromic, so it must be hyperbolic; and so on.

11. The Isometric Circle.—In an analytic transformation $z' = f(z)$, a lineal element $dz = z_2 - z_1$ connecting two points in the infinitesimal neighborhood of a point z is transformed

into the lineal element dz' in the neighborhood of z' . We have $dz' = f'(z)dz$; hence, the length of the element is multiplied by $|f'(z)|$, and the element is rotated through an angle $\arg f'(z)$. For the linear transformation

$$z' = T(z) = \frac{az + b}{cz + d}, \quad ad - bc = 1, \quad (1)$$

we have the following theorem:

THEOREM 16.—When the transformation (1) is applied, infinitesimal lengths in the neighborhood of a point z are multiplied by $|cz + d|^{-2}$; infinitesimal areas in the neighborhood of z are multiplied by $|cz + d|^{-4}$.

For we have

$$\frac{dz'}{dz} = T'(z) = \frac{1}{(cz + d)^2}, \quad (40)$$

whence, lengths are multiplied by $|T'(z)|$, or $|cz + d|^{-2}$. An infinitesimal region is carried into a similar region with corresponding lengths multiplied by $|T'(z)|$; hence, the area is multiplied by $|T'(z)|^2$, or $|cz + d|^{-4}$.

We get alternative forms for $T'(z)$ from (24) and (37). Differentiating (24) and simplifying, we obtain the first of the following results; inverting and differentiating, we find the second. For the transformation with two fixed points,

$$\frac{dz'}{dz} = T'(z) = K \left(\frac{z' - \xi_2}{z - \xi_2} \right)^2 = \frac{1}{K} \left(\frac{z' - \xi_1}{z - \xi_1} \right)^2. \quad (41)$$

For the parabolic transformation (37), we have

$$T'(z) = \left(\frac{z' - \xi}{z - \xi} \right)^2, \quad (42)$$

the same as (41) with the value $K = 1$.

From (41), we have

$$T'(\xi_1) = K, \quad T'(\xi_2) = \frac{1}{K}. \quad (43)$$

At ξ_1 which is fixed, $dz' = Kdz$. For the hyperbolic transformation, $K = A$, the infinitesimal neighborhood of ξ_1 undergoes a stretching from ξ_1 ; for the elliptic transformation, $K = e^{i\theta}$, there is a rotation about ξ_1 through the angle θ ; for the loxodromic transformation, $K = Ae^{i\theta}$, there is a combination of stretching and rotation. Analogous remarks apply to the neighborhood of

ξ_2 . Since $\frac{1}{K} = \frac{1}{A}e^{-i\theta}$, the stretching and rotation are in the opposite sense.

In the parabolic transformation, we find, on substituting $\xi = (a - d)/2c$ into (40), that $T'(\xi) = 1$. The infinitesimal neighborhood of ξ is unaltered.

Lengths and areas are unaltered in magnitude if, and only if, $|cz + d| = 1$. If $c \neq 0$, the locus of z is a circle. Writing $z + \frac{d}{c} = \frac{1}{c}$, we see that the center is $-\frac{d}{c}$, the radius is $\frac{1}{|c|}$.

DEFINITION.—The circle I ,

$$I: |cz + d| = 1, \quad c \neq 0, \quad (44)$$

which is the complete locus of points in the neighborhood of which lengths and areas are unaltered in magnitude by the transformation (1), is called the isometric circle of the transformation.

The isometric circle will play a fundamental part in many of our later developments. In this section we shall investigate some of its properties.

We note, first, that if $c = 0$, so that ∞ is a fixed point, there is no unique circle with the property of the isometric circle. The derivative $T'(z)$ is constant and equal to K (Equations (31) and (39')). Either $|K| \neq 1$, and all lengths are altered in magnitude; or $|K| = 1$, and all lengths are unaltered. The latter case comprises the rigid motions—the rotations (31) and translations (39').

THEOREM 17.—Lengths and areas within the isometric circle are increased in magnitude, and lengths and areas without the isometric circle are decreased in magnitude, by the transformation.

For, if z is within I , $\left| z + \frac{d}{c} \right| < \frac{1}{|c|}$, or $|cz + d| < 1$, and $|T'(z)| > 1$. A length or area within I is thus magnified in all its parts. Similarly, if z is without I , $|T'(z)| < 1$; and a length or an area without I is diminished in all its parts.

THEOREM 18.—A transformation carries its isometric circle into the isometric circle of the inverse transformation.

The inverse transformation, $z' = (-dz + b)/(cz - a)$, has the isometric circle

$$I': |cz - a| = 1. \quad (45)$$

Its center is a/c , its radius $1/|c|$. Now T carries I into a circle I_0 without alteration of lengths in the neighborhood of any point,

hence T^{-1} carries I_0 back to I without alteration. But I' is the complete locus of points in the neighborhood of which T effects no change of length; hence, I_0 coincides with I' .

(a) *Geometric Interpretation of the Transformation.*—The transformation T carries I into I' (Fig. 9) without alteration of any arc. Let a point P on I be carried into P' . Then, if I be set down upon I' so that P coincides with P' , with proper orientation, corresponding points will coincide. Any sequence of an even number of inversions which will effect the proper transformation on I will be equivalent to T (Theorem 7).

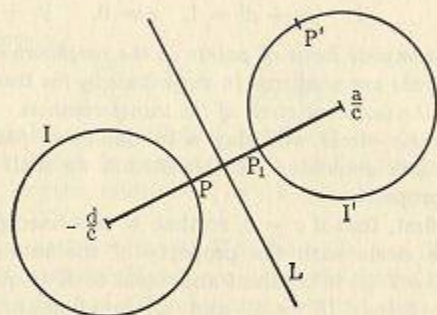


FIG. 9.

As a point moves from P counterclockwise around I , suppose that the corresponding point moves from P' counterclockwise around I' . Then, I can be carried into I' by a rigid motion so that corresponding points coincide. But ∞ is fixed for a rigid motion, so $c = 0$; hence, this case is impossible. Consequently, as a point moves counterclockwise around I the corresponding point moves clockwise around I' . The circle I must be turned over before being applied to I' .

An inversion in I , leaving the points of I invariant, followed by a reflection in L , the perpendicular bisector of the line segment joining the centers, carries I into I' with the desired change of order. P is carried into a point P_1 . A rotation with a/c fixed will carry P_1 into P' . The two inversions together with the rotation are equivalent to T .

Since a rotation is equivalent to two reflections (Sec. 5(b)), four inversions at most are adequate for the representation of the

transformation. If P' coincides with P_1 , two inversions are sufficient.

Several alternative geometric transformations are possible. Thus, instead of inverting in I and then reflecting in L we may reflect in L and then invert in I' . Or we may rotate about $-d/c$ at the start; and so on.

The preceding construction fails if I and I' coincide, for then L is not defined. In this case $a = -d$, or $a + d = 0$; and T is an elliptic transformation of period two (Sec. 8). P' lies on I . An inversion in I followed by a reflection in L , the line joining $-d/c$ to the midpoint of the arc PP' is equivalent to T .

(b) *The Types of Transformations.*—The distance between the centers of I and I' is $\left| \frac{a}{c} + \frac{d}{c} \right|$; the sum of the radii is $2/|c|$. The circles will intersect, touch, or be totally exterior according as $|a + d|$ is less than, equal to, or greater than 2. Hence, applying Theorem 15, if T is hyperbolic, the isometric circles of T and T^{-1} are external; if T is elliptic, they intersect; if T is parabolic, they are tangent. If T is loxodromic, $|a + d|$ may have any value other than zero, and the isometric circles may have any relation to one another other than coincidence.

A distinction between the loxodromic and the three non-loxodromic transformations appears when we study the geometrical transformations which are equivalent to the transformation. Let P, P_1, P' (arranged as in Fig. 9) have the coordinates z, z_1, z' .

Since $z, -d/c$, and a/c lie on a line, $z + \frac{d}{c}$ and $\frac{a}{c} + \frac{d}{c}$ have the same argument, the moduli being $1/|c|$ and $|a + d|/|c|$, respectively, hence,

$$z + \frac{d}{c} = \frac{a + d}{|a + d|c}.$$

Similarly,

$$z_1 - \frac{a}{c} = -\frac{a + d}{|a + d|c}.$$

Making the transformation T , using (22'),

$$z' - \frac{a}{c} = -\frac{1}{c^2(z + d/c)} = -\frac{|a + d|}{(a + d)c} = -\frac{\bar{a} + \bar{d}}{|a + d|c}.$$

From these equations we see that z' coincides with z_1 if, and only if, $\bar{a} + \bar{d} = a + d$; that is, if $a + d$ is real. The transformation is then non-loxodromic.

If the transformation is loxodromic, writing $a + d = re^{i\varphi}$, $\varphi \neq n\pi$, we have

$$z' - \frac{a}{c} = \frac{\bar{a} + \bar{d}}{a + d} \left(z_1 - \frac{a}{c} \right) = e^{-2i\varphi} \left(z_1 - \frac{a}{c} \right).$$

To carry z_1 to z' , there is a rotation about a/c through the angle -2φ . We have the result:

THEOREM 19.—If the transformation is hyperbolic, elliptic, or parabolic, it is equivalent to an inversion in I followed by a reflection in L ; if it is loxodromic there is in addition a rotation about the center of I' through the angle $-2 \arg (a + d)$.

Consider now the fixed points. Since $T'(\xi_1) = K$, $T'(\xi_2) = 1/K$ (Equation (43)), we have, if $|K| \neq 1$, increase of lengths at one fixed point and decrease at the other; if $|K| = 1$, there is no alteration. Hence, for the hyperbolic and loxodromic transformations one fixed point is within I , the other without; for the elliptic transformation both fixed points, and for the parabolic transformation the single fixed point, are on I . Identical statements are true of I' for similar reasons.

In the elliptic transformation I and I' intersect and L is the common chord. The points of intersection are fixed for both the inversion in I and the reflection in L ; hence, they are the fixed points. We found that the lineal elements issuing from the fixed point are rotated through an angle θ , where $K = e^{i\theta}$. Since an arc of I issuing from the fixed point is transformed into an arc of I' issuing from the point, it follows that I and I' intersect at the angle θ .

If $a + d = 0$, so that I and I' coincide, the line L is the line joining the fixed points, which are then at the ends of a diameter.

In the parabolic transformation, L is the common tangent to I and I' at their point of tangency. The point of tangency is then the fixed point.

(c) *The Fixed Circles.*—We consider now the non-loxodromic transformations. Each such transformation has a one-parameter family of fixed circles, including, as we found in Sec. 10, the line joining the centers of I and I' . The family of fixed circles is easily constructed. It consists of the circles with centers on L orthogonal to I . For, being orthogonal to I , such a circle is transformed into itself by an inversion in I ; and a reflection in L , a diameter, transforms it again into itself. Each fixed circle is also orthogonal to I' from symmetry.

THEOREM 20.—In a non-loxodromic transformation the isometric circle is orthogonal to the fixed circles.

For use later we shall prove the following theorem:

THEOREM 21.—Let Q be a fixed circle of a non-loxodromic transformation and I its isometric circle. Let h be the distance of a point z from q , the center of Q , and h' be the distance of the transformed point z' from q ; then

$h' = h$, if z is on I or Q ;

$h' < h$, if z is within both I and Q , or without both;

$h' > h$, if z is within either I or Q , and without the other.

An inversion in I carries z to a point z_1 which is carried to z' by a reflection in L (Fig. 10). Obviously, the reflection does not alter distances from q . The proposition hinges, then, on what happens when z is inverted in I .

The distances of a point and its inverse from the center of a circle orthogonal to the circle of inversion is clearly independent of the orientation of the circles, and their relative magnitudes are independent of the scale used; hence, it will suffice to take for I the unit circle $z\bar{z} = 1$ and to take q on the real axis. The equation of Q is $(z - q)(\bar{z} - q) = r^2$, where, for orthogonality, $r^2 + 1 = q^2$; whence,

$$z\bar{z} - q(z + \bar{z}) + 1 = 0.$$

The expression in the first member of this equation is positive for points without Q and negative for points within. Now,

$$h^2 = (z - q)(\bar{z} - q) = z\bar{z} - q(z + \bar{z}) + 1 + r^2,$$

and, since $z_1 = 1/\bar{z}$,

$$h'^2 = z_1\bar{z}_1 - q(z_1 + \bar{z}_1) + 1 + r^2 = \frac{1 - q(\bar{z} + z) + z\bar{z}}{z\bar{z}} + r^2;$$

whence,

$$h^2 - h'^2 = \frac{[z\bar{z} - 1][z\bar{z} - q(z + \bar{z}) + 1]}{z\bar{z}}.$$

The theorem follows immediately from this equation. If z is within both circles or without both circles, the factors in the

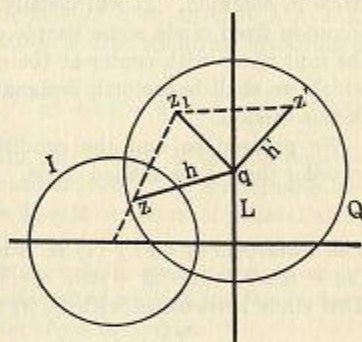


FIG. 10.

numerator of the second member are both negative or both positive, and $h' < h$; if z is within one circle and without the other, the factors differ in sign, and $h' > h$; if z is on one circle, one factor is zero, and $h' = h$.

The following theorem relative to the fixed straight line is easily seen to hold:

THEOREM 22.—*In a non-loxodromic transformation let k and k' be the distances of z and z' , respectively, from the fixed straight line M ; then $k' = k$ if z is on I or on M ; otherwise, $k' > k$ if z is within I and $k' < k$ if z is without I .*

12. The Unit Circle.—We shall, subsequently, have much to do with sets of linear transformations which have one fixed circle in common. It will usually be convenient to take as the common fixed circle some simple circle such as the real axis or the unit circle with center at the origin. It is this latter circle, which we shall henceforth designate by Q_0 , that we shall study in this section.

We proceed to find the conditions on the constants in (1) in order that Q_0 be a fixed circle. The equation of Q_0 is

$$z\bar{z} - 1 = 0. \quad (46)$$

The transform of Q_0 by (1) is, from Equation (19),
 $(d\bar{d} - c\bar{c})z'\bar{z}' + (-b\bar{d} + a\bar{c})z' + (-b\bar{d} + a\bar{c})\bar{z}' + b\bar{b} - a\bar{a} = 0$.
 This circle is identical with Q_0 if, and only if,

$$-b\bar{d} + a\bar{c} = 0, \quad -b\bar{d} + a\bar{c} = 0. \quad (a)$$

$$d\bar{d} - c\bar{c} = a\bar{a} - b\bar{b} \neq 0. \quad (b)$$

Each equation in (a) is a consequence of the other; from the second,

$$\frac{b}{\bar{c}} = \frac{a}{\bar{d}} = \lambda,$$

say; then

$$b = \lambda\bar{c}, \quad a = \lambda\bar{d}.$$

Substituting in (b)

$$d\bar{d} - c\bar{c} = \lambda\bar{\lambda}(d\bar{d} - c\bar{c}) \neq 0,$$

whence, $\lambda\bar{\lambda} = 1$.

From $ad - bc = 1$, we have

$$\lambda(d\bar{d} - c\bar{c}) = 1;$$

hence, λ is real, so $\lambda = \pm 1$. The sign of λ depends upon the sign of $d\bar{d} - c\bar{c}$. If the interior of Q_0 is transformed into its interior, the point $-d/c$ which is carried to ∞ is outside the

circle and $|-d/c| > 1$; so $d\bar{d} - c\bar{c} > 0$ and $\lambda = 1$. We have, then,

$$b = \bar{c}, \quad a = \bar{d}, \quad d = \bar{a}.$$

These values obviously satisfy the conditions (a) and (b). We have the following result:

THEOREM 23.—*The most general linear transformation carrying Q_0 into itself and carrying the interior of Q_0 into itself is the transformation*

$$z' = \frac{az + \bar{c}}{cz + \bar{a}}, \quad a\bar{a} - c\bar{c} = 1. \quad (47)$$

The most general linear transformation carrying Q_0 into itself and carrying the interior of Q_0 into its exterior is found similarly. Then, $-d/c$ is within Q_0 and $d\bar{d} - c\bar{c} < 0$; so that $\lambda = -1$. The most general transformation is the resulting loxodromic transformation

$$z' = \frac{az - \bar{c}}{cz - \bar{a}}, \quad c\bar{c} - a\bar{a} = 1. \quad (48)$$

The transformation (47) maps the interior of Q_0 in a one-to-one and directly conformal manner on itself. It is a remarkable fact, which we shall now prove, that it is the most general such transformation. We first prove the following proposition:

The most general transformation which maps the interior of Q_0 in a one-to-one and directly conformal manner on itself and which leaves the origin fixed is a rotation about the origin.

Let $z' = f(z)$ be such a transformation. Owing to the conformality $f(z)$ is analytic in Q_0 . Further, $|f(z)| < 1$ when $|z| < 1$, since an interior point is carried into an interior point. Since $z' = 0$ when $z = 0$, $f(z)$ has a zero at the origin; hence $f(z)/z$ is analytic in Q_0 .

Consider now $|f(z)/z|$ in a circle Q' with center at the origin and radius $r < 1$. Since the absolute value of a function which is analytic in and on the boundary of a region takes on its maximum value on the boundary, we have, since $|z| = r$ on Q' ,

$$\left| \frac{f(z)}{z} \right| < \frac{1}{r} \text{ in } Q'.$$

Since r may be taken as near to 1 as we like, we have

$$\left| \frac{z'}{z} \right| = \left| \frac{f(z)}{z} \right| \leq 1 \text{ in } Q_0.$$

Considering the inverse function, $z = \varphi(z')$, we have by the same reasoning $|z/z'| \leq 1$ in Q_0 . Consequently, $|z'/z| = 1$ in Q_0 ,

whence $z'/z = e^{i\alpha}$. But, if the absolute value of an analytic function is constant, so also is its argument; hence α is constant. We have, thus,

$$z' = e^{i\alpha}z,$$

a rotation about the origin.

We shall now remove the restriction that the origin be fixed. Let $z' = f(z)$ map the interior of Q_0 on itself in a one-to-one and directly conformal manner, and let $f(0) = z_0$. Let $z' = S(z)$ be a linear transformation of the form (47) such that $S(0) = z_0$. (It is easy to determine the constants of (47) so that $\bar{c}/\bar{a} = z_0 < 1$.) If we make the transformation f and then make S^{-1} , the interior of Q_0 is carried into itself and the origin is fixed. Hence, $S^{-1}f = U$, a rotation; and $f = SU$. We thus have a linear transformation. Since it carries the interior of Q_0 into itself, it is of the form (47).

THEOREM 24.—*The most general transformation which maps the interior of Q_0 in a one-to-one and directly conformal manner on itself is the linear transformation (47).*

The proof of the following more general theorem is now easily made.

THEOREM 25.—*The most general transformation which maps the interior or exterior of one circle in a one-to-one and directly conformal manner upon the interior or exterior of another circle is a linear transformation.*

Let $z' = f(z)$ carry the interior or exterior of Q_1 into the interior or exterior of Q_2 in the manner stated. Let S_1 and S_2 be linear transformations carrying Q_1 and Q_2 , respectively, into Q_0 , the interior or exterior of each which is involved in the mapping being carried into the interior of Q_0 . Then, the sequence of transformations, S_1^{-1} , followed by f , followed by S_2 , carries the interior of Q_0 into itself, and is equivalent to a linear transformation T of the form (47).

$$S_2 f S_1^{-1} = T, \text{ or } f = S_2^{-1} T S_1.$$

The transformation is thus a linear transformation.

CHAPTER II

GROUPS OF LINEAR TRANSFORMATIONS

13. Definition of a Group. Examples.—The automorphic function depends for its definition on a set of linear transformations called a "group." In the present chapter we shall make a study of groups of linear transformations, after which we shall be in a position to pass to the definition of the automorphic function and to a study of its properties.

DEFINITION.—*A set of transformations, finite or infinite in number, is said to form a group if,*

(a) *the inverse of each transformation of the set is a transformation of the set;*

(b) *the succession of any two transformations of the set is a transformation of the set.*

The definition applies to all kinds of transformations, but we shall be concerned only with sets of linear transformations. The two group properties, expressed in symbolic notation, are: (a) if T is any transformation of the set so also is T^{-1} ; (b) if S is a transformation of the set, not necessarily different from T , so also is ST . It follows by a repeated application of (b) that the transformation equivalent to performing any sequence of transformations of a group belongs to the group. In particular, all positive and negative integral powers of a transformation T of the group belong to the group. Also $T^{-1}T (=1)$ belongs to the group; that is, *every group contains the identical transformation, $z' = z$.*

Given a set of linear transformations, we may test whether or not it constitutes a group by applying (a) and (b) to the transformations of which it is composed. There are, however, certain cases in which the group properties obviously hold. For example, if the set consists of *all* linear transformations which leave some configuration F in the z -plane invariant, then the set is a group; for, clearly, the inverse of any transformation or the successive performance of any two transformations will leave F invariant and, being themselves linear transformations, will

then belong to the set. Thus, all linear transformations with a common fixed point constitute a group. All linear transformations of the form (47), Sec. 12, leaving Q_0 and its interior invariant, form a group. The set of all linear transformations which carry a given regular polygon into itself, consisting of certain rotations about its center, form a group.

Similarly, the set of all linear transformations which leave invariant some function of z constitute a group. For example, all linear transformations $z' = T(z)$ such that $\sin z' = \sin z$ form a group. Such transformations as $z' = z + 2\pi$, $z' = z + 4\pi$, $z' = \pi - z$, etc., belong to this group. It is by virtue of this property, as we shall see later, that $\sin z$ is called an "automorphic function."

Given a set of linear transformations $T_1, T_2, \dots, T_n$, we may form a group containing them in the following way: Let the set contain the given transformations, their inverses, and the transformations formed by combining the given transformations and their inverses into products in all possible ways. Then it is easily seen that the inverse of any transformation or the product of any two is itself some combination of the given transformations and their inverses and, consequently, is included in the set. Hence, the whole set forms a group. The group is said to be "generated" by the transformations $T_1, T_2, \dots, T_n$, and the transformations are called "generating transformations" of the group.

Examples.—The following are a few examples of well-known groups, some of which will be discussed later.

1. *A Group of Rotations about the Origin.*—The m transformations, $z' = z, e^{2\pi i/m}z, e^{4\pi i/m}z, \dots, e^{2(m-1)\pi i/m}z$, form a group. They are the rotations about the origin through multiples of the angle $2\pi/m$. The group is generated by the transformation $z' = e^{2\pi i/m}z$.

2. *The Group of Anharmonic Ratios.*—The six transformations

$$z' = z, \quad \frac{1}{z}, \quad 1 - z, \quad \frac{1}{1 - z}, \quad \frac{z - 1}{z}, \quad \frac{z}{z - 1}$$

form a group. It can be verified by forming the inverses and by combining the transformations that both group properties are satisfied. The group is so named for the reason that if z is any one of the anharmonic ratios of four points on a line, the six anharmonic ratios are given by the transformations of the group.

3. *The Group of the Simply Periodic Functions.*—The set $z' = z + m\omega$, where ω is a constant different from zero, and m is any positive or negative integer or zero, forms a group. The group is generated by the transformation $z' = z + \omega$.

4. *The Group of the Doubly Periodic Functions.*—The set $z' = z + m\omega + m'\omega'$, where ω and ω' are constants different from zero and the ratio ω'/ω is not real, and where m and m' are any positive or negative integers or zero, forms a group. It is generated by the transformations $z' = z + \omega, z' = z + \omega'$.

The restrictions on ω and ω' are not necessary for establishing the group properties.

5. *The Modular Group.*—The infinite set of transformations $z' = (az + b)/(cz + d)$, where a, b, c, d are real integers such that $ad - bc = 1$, constitutes a group. For, a reference to Equations (3) and (6) of Sec. 1 shows that the inverse of such a transformation and, also, the product of two such transformations are transformations with integral coefficients and of unit determinants. Since the coefficients are real, each transformation carries the real axis into itself.

6. *The Group of Picard.*—The set of transformations $z' = (az + b)/(cz + d)$, where a, b, c, d are either real or complex integers (i.e. of the form $m + ni$, where m and n are real integers) such that $ad - bc = 1$, constitutes a group. The proof is as in the preceding case.

7. *A Group Allied to Q_0 .*—In a similar manner, the transformations $z' = (az + \bar{c})/(cz + \bar{a})$, where a and c are real or complex integers such that $a\bar{a} - c\bar{c} = 1$, form a group. The transformations of this group (Theorem 23, Sec. 12) have Q_0 as a fixed circle and carry the interior of Q_0 into itself.

14. *Properly Discontinuous Groups.*—If we compare the group of the simply periodic functions, $z' = z + m\omega$, with the group of all translations, $z' = z + b$, where b is any constant, we observe the following difference: In the former case there is no transform of a point z within the distance $|\omega|$ of z ; in the latter group we get transforms of z as near to z as we like by taking b small enough. These two groups bring out an essential distinction.

DEFINITION.—A group is called properly discontinuous in the z -plane if there exists a point z_0 and a region S enclosing z_0 such that all transformations of the group, other than the identical transformation, carry z_0 outside S .

The automorphic functions are founded on the properly discontinuous groups, and these only will appear in our subsequent study. The groups whose transformations contain continuously varying parameters, which have given rise to so many and so profound researches, play no part in the theory to which this book is devoted and will not be considered further.

A group is said to contain infinitesimal transformations if there is, for some region A and any given $\epsilon > 0$, a transformation¹ $z' = (az + b)/(cz + d)$, $ad - bc = 1$, such that for all points z of A we have $|z' - z| < \epsilon$. It is found without difficulty that a necessary and sufficient condition for this

¹ Other than the identical transformation.

is that there be transformations for which c , $d - a$, and b are all arbitrarily small (but not all zero, for then we have the identical transformation).

Not all groups which are free of infinitesimal transformations are properly discontinuous. The group of Picard, for example, does not contain infinitesimal transformations, since c , $d - a$, and b are complex integers and cannot be made arbitrarily small without being all zero. It can be shown, however, that the points into which any point is carried are everywhere dense in the whole z -plane. Such a group, that is, one which does not contain infinitesimal transformations and yet which is not properly discontinuous in the z -plane, is called "improperly discontinuous" in the z -plane.

15. Transforming a Group.—From a given group of linear transformations infinitely many other groups can be derived by applying linear transformations to the plane in which z and its transforms are represented. Let T be any transformation of the given group, and let T carry z into z' . Let a transformation G be applied, z and z' being transformed into z_1 and z_1' , respectively. Then z_1 is carried into z_1' by the transformation S where

$$S = GTG^{-1}; \quad (1)$$

for

$$GTG^{-1}(z_1) = GT(z) = G(z') = z_1'.$$

Let all the transformations of the original group be altered in this manner, so that to each T of the group there corresponds an S given by (1). We shall show that the new set of transformations forms a group. We have $S^{-1} = (GTG^{-1})^{-1} = GT^{-1}G^{-1}$, which belongs to the set since T^{-1} belongs to the original group. If $S_1 = GT_1G^{-1}$ is a second transformation of the set, $SS_1 = GTG^{-1}GT_1G^{-1} = GTT_1G^{-1}$, and SS_1 belongs to the set since TT_1 belongs to the original group. Thus, both group properties are satisfied. Two groups whose transformations can be made to correspond in a one-to-one manner, as the S and T transformations are paired by virtue of (1), so that the product of any number of transformations of one group corresponds to the analogous product of the corresponding transformations of the other, are said to be "isomorphic."

It will often facilitate the study of a group to transform it in the manner indicated. For example, an important point can be carried to ∞ , or an important circle can be carried into the real axis or the unit circle Q_0 . Having found how figures are transformed by the new group, we can then carry the results back to the old by applying G^{-1} . For if S carries a figure F into F' , T carries $G^{-1}(F)$ into $G^{-1}(F')$, as we see at once from the equation $T = G^{-1}SG$.

It should be mentioned that the transformations S and T of (1) are of the same type, whatever G may be. Let

$$T = \frac{az + b}{cz + d}, \quad G = \frac{\alpha z + \beta}{\gamma z + \delta}, \quad \begin{aligned} ad - bc &= 1, \\ \alpha\delta - \beta\gamma &= 1, \end{aligned}$$

and form the product in (1), using the equations (3) and (6) of Sec. 1,

$$S = \frac{(-\alpha\delta a + \alpha\gamma b - \beta\delta c + \beta\gamma d)z + \alpha\beta a - \alpha^2 b + \beta^2 c - \alpha\beta d}{(-\gamma\delta a + \gamma^2 b - \delta^2 c + \gamma\delta d)z + \beta\gamma a - \alpha\gamma b + \beta\delta c - \alpha\delta d} \quad (2)$$

the determinant of S being 1. Then,

$$\begin{aligned} &(-\alpha\delta a + \alpha\gamma b - \beta\delta c + \beta\gamma d) + (\beta\gamma a - \alpha\gamma b + \beta\delta c - \alpha\delta d) \\ &= -(a + d). \end{aligned}$$

It follows that K has the same value for S as for T (Equation (26), Sec. 6).

16. The Fundamental Region.—Before proceeding to the study of the general properly discontinuous group it is desirable to introduce the important concept of the fundamental region.

DEFINITION.—Two configurations (points, curves, regions, etc.) are said to be congruent with respect to a group if there is a transformation of the group other than the identical transformation, which carries one configuration into the other.

DEFINITION.—A region, connected or not, no two of whose points are congruent with respect to a given group, and such that the neighborhood of any point on the boundary contains points congruent to points in the given region, is called a fundamental region for the group.

The accompanying figures show fundamental regions for certain simple groups. The reader is probably already familiar with some of them.

For the group of rotations about a point through multiples of an angle θ , which is a submultiple of 2π , we draw two half lines from the fixed point forming an angle θ . The region R_0 within the angle is a fundamental region. In Fig. 11, R_0 is a fundamental region for the group $z' = e^{i2\pi/c}z$. The neighborhood of any point on the boundary of R_0 contains points which can be carried into the interior of R_0 by a rotation through the angle $\pm 2\pi/c$.

In Fig. 12, R_0 , whose construction is evident from the figure, is a fundamental region, or period strip, for the group of the simply periodic functions, $z' = z + m\omega$.

In Fig. 13, R_0 is a fundamental region, or period parallelogram, for the group of the doubly periodic functions $z' = z + m\omega + m'\omega'$.

In Fig. 14, R_0 , which is bounded by circles with centers at the origin and with radii 1 and A , is a fundamental region for the group of stretchings from the origin, $z' = Az$.

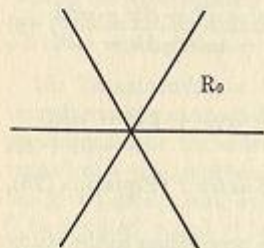


FIG. 11.

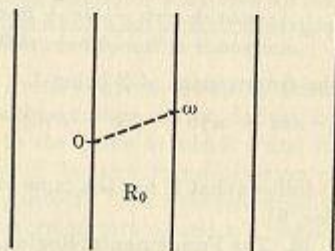


FIG. 12.

Attention may be called to certain properties that are common to the four fundamental regions constructed in the figures and which we shall find to be more or less generally true—to what extent will appear from later analysis—of the fundamental regions we shall use for less simple groups. We note first that the

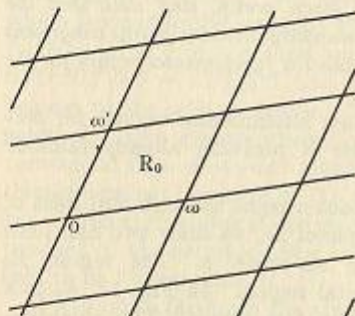


FIG. 13.

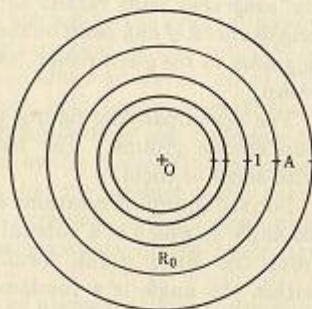


FIG. 14.

boundaries of R_0 in each case consist of congruent curves. In Figs. 11, 12, 14, each of the two boundaries can be carried into the other by a transformation of the group. In Fig. 13, the lower boundary can be carried into the upper by the translation $z' = z + \omega$, and the left into the right by $z' = z + \omega'$.

Further, the transformations connecting congruent boundaries are generating transformations of the group. The two translations just mentioned generate the group of doubly periodic functions. In Fig. 11, all transformations are formed by successive applications of the rotation $z' = e^{2\pi i/z}z$, which carries one boundary into the other. The like fact is true of the other examples.

We note that we can add to the open region R_0 one, but not both, of two congruent boundaries without getting two congruent points in the region. But R_0 must remain in part an open region.

The region R_0 and the regions congruent to it, some of which are shown in the figures, form a set of adjacent, non-overlapping regions covering practically the whole plane. The origin in Fig. 14, however, is not in any region congruent to R_0 .

The angle at the vertex of R_0 in Fig. 11 is a submultiple of 2π . The sum of the angles at the four congruent vertices of R_0 in Fig. 13 is equal to 2π . These facts will reappear, suitably generalized.

It is clear that the fundamental region is in no wise unique. Any region congruent to R_0 will serve as a region. Furthermore, we can replace any part of R_0 by a congruent part and still have a fundamental region. Thus, we can subtract a part at one boundary and add a congruent part at another. In this way the character of the bounding curve can be altered freely.

THEOREM 1.—If no two points of a region are congruent, the transforms of the region by two distinct transformations of the group do not overlap.

Let A be a region containing no two congruent points. Suppose that two transformations of the group, S and T , carry A into two overlapping regions. Any point z_0 in the common part is the transform by S of a point z_1 of A and the transform by T of a point z_2 of A . If z_1 and z_2 are different for any z_0 , then z_1 and z_2 are congruent points of A , which is impossible. If z_1 and z_2 coincide for every z_0 in the common part, S and T are the same transformation (Theorem 7, Sec. 3).

Since a fundamental region contains no two congruent points, we can state the following useful corollary:

COROLLARY.—The transforms of a fundamental region by two distinct transformations of the group do not overlap.

17. The Isometric Circles of a Group.—We shall now investigate the properties of the most general properly discontinuous

group. For such a group there exists, by hypothesis, at least one point z_0 such that there are no transforms of z_0 in a suitably small region about z_0 . Let G be a transformation carrying z_0 into ∞ ; and let the group be transformed by G as explained in Sec. 15. It is this transformed group which we shall study.

There is no point congruent to ∞ outside or on a circle Q_ρ with a given center and with radius ρ suitably large. In particular, ∞ is not a fixed point for any transformation of the group. Hence, in any transformation $T = (az + b)/(cz + d)$ we have $c \neq 0$, except in the case of the identical transformation. The center of an isometric circle is congruent to ∞ ($-d/c$ is carried to ∞ by T); hence, the centers of all isometric circles lie within Q_ρ .

(a) *The Isometric Circle of the Product of Two Transformations.* Certain relations between the isometric circles of two transformations and the isometric circle of their product will be of use now and subsequently. Consider any two transformations

$$T = \frac{az + b}{cz + d}, \quad S = \frac{\alpha z + \beta}{\gamma z + \delta}, \quad \begin{aligned} ad - bc &= 1, c \neq 0, \\ \alpha\delta - \beta\gamma &= 1, \gamma \neq 0. \end{aligned}$$

Then,

$$ST = \frac{(\alpha a + \beta c)z + \alpha b + \beta d}{(\gamma a + \delta c)z + \gamma b + \delta d} \quad (3)$$

In what follows we assume that $S \neq T^{-1}$ so that ST is not the identical transformation; then, the isometric circle of ST is

$$|(\gamma a + \delta c)z + \gamma b + \delta d| = 1. \quad (4)$$

Represent by $I_s, I_s', I_t, I_t', I_{st}$ the isometric circles of S, S^{-1}, T, T^{-1}, ST , respectively; by $g_s, g_s', g_t, g_t', g_{st}$ their respective centers; and by r_s, r_t, r_{st} their radii. We have

$$\begin{aligned} g_s &= -\frac{\delta}{\gamma}, & g_s' &= \frac{\alpha}{\gamma}, & g_t &= -\frac{d}{c}, & g_t' &= \frac{a}{c}, & g_{st} &= -\frac{(\gamma b + \delta d)}{(\gamma a + \delta c)}, \\ r_s &= \frac{1}{|\gamma|}, & r_t &= \frac{1}{|c|}, & r_{st} &= \frac{1}{|\gamma a + \delta c|}. \end{aligned} \quad (5)$$

From these values, we derive the following relations:

$$r_{st} = \frac{1}{|\gamma a + \delta c|} = \frac{1}{|\gamma c| \cdot \left| \frac{a}{c} + \frac{\delta}{\gamma} \right|} = \frac{r_s r_t}{|g_t' - g_s|} \quad (6)$$

and

$$g_{st} - g_t = \frac{\gamma b + \delta d}{\gamma a + \delta c} + \frac{d}{c} = \frac{\gamma}{c(\gamma a + \delta c)}, \quad (7)$$

whence,

$$|g_{st} - g_t| = \frac{r_{st} r_t}{r_s} = \frac{r_t^2}{|g_t' - g_s|}. \quad (8)$$

(b) *The Arrangement of the Isometric Circles.*—By means of the preceding equations we can derive certain simple facts concerning the isometric circles of the transformations of the group.

The radii of the isometric circles are bounded.

Let $T (\neq 1)$ be any transformation of the group and $S (\neq 1)$ a transformation of the group different from T^{-1} . Then, from (8),

$$r_t^2 = |g_{st} - g_t| \cdot |g_t' - g_s|.$$

But each factor in the second member, being the distance between points of Q_ρ , is less than 2ρ ; hence,

$$r_t^2 < 4\rho^2, \quad r_t < 2\rho. \quad (9)$$

The number of isometric circles with radii exceeding a given positive quantity is finite.

Let I_s and I_t' be any two different isometric circles with radii greater than k , a positive quantity. Then ST is not the identical transformation, and, from (6),

$$|g_t' - g_s| = \frac{r_s r_t}{r_{st}} > \frac{k^2}{2\rho}. \quad (10)$$

The distance between the centers of two isometric circles with radii exceeding k has thus a positive lower bound. Since the centers of all such circles lie in the circle Q_ρ , their number must be finite.

It follows from this fact that the transformations of the group are denumerable.

Another consequence may be stated in the following manner:

Given any infinite sequence of distinct isometric circles $I_1, I_2, I_3, \dots$, of transformations of the group, the radii being $r_1, r_2, r_3, \dots$, then $\lim_{n \rightarrow \infty} r_n = 0$.

18. The Limit Points of a Group.—In this section and the remaining sections of the present chapter we suppose that no transformation of the group has a fixed point at infinity, so that the isometric circles exist for all transformations except the identical transformation; and that there are no points congruent to infinity in the neighborhood of infinity. This assumption involves no essential restriction since, as we have already noted, any properly discontinuous group can be transformed into one with the properties mentioned.

Consider the centers of the isometric circles. If the group contains an infinite number of transformations, the centers are

infinite in number and, hence, have one or more cluster points. We lay down the following definitions:

DEFINITIONS.—A cluster point of the centers of the isometric circles of the transformations of a group is called a limit point of the group.

A point which is not a limit point is called an ordinary point.

It is clear that all limit points lie within or on the circle Q_p of Sec. 17, since the centers of all isometric circles lie within that circle. If the group contains only a finite number of transformations, there are, of course, no limit points.

THEOREM 2.—In the neighborhood of a limit point P there is an infinite number of distinct points congruent to any point of the plane, with, at most, the exception of P itself and of one other point.

Since only a finite number of isometric circles have radii exceeding a given positive quantity, there are isometric circles of arbitrarily small radius in the neighborhood of P . Let Q be a small circle about P and let $I_1', I_2', \dots$ be an infinite sequence of isometric circles contained in Q , where the center g_n approaches P as n becomes infinite. Let these be the transforms of the isometric circles $I_1, I_2, \dots$, and let S_n be the transformation carrying I_n into I_n' .

The centers g_n of I_n have at least one cluster point. Suppose first that there is such a point P' distinct from P . It will suffice to show that for any point P_1 distinct from P and P' there is a congruent point in Q distinct from P ; for, by decreasing the region Q , we then have an infinite number of congruent points. Let I_n be near P' and of small enough radius that I_n encloses neither P_1 nor P . Then, since P_1 is outside I_n , $S_n(P_1)$ is inside I_n' and in Q . If $S_n(P_1)$ is different from P , the proposition is established. If $S_n(P_1)$ coincides with P , then P is not a fixed point for S_n and $S_n^2(P_1)$, or $S_n(P)$, is in Q and different from P .

There remains the case that the only limit point of the centers g_n is P itself. Let A and A' be any two points distinct from one another and from P . A and A' are outside an infinite number of circles I_n . For these circles the congruent points, $A_n = S_n(A)$ and $A_n' = S_n(A')$ are in I_n' and in Q . At least one of the points A_n and A_n' is distinct from P . It follows that at least one of the points A and A' has an infinite number of congruent points in Q which are distinct from P . Hence, there is not more than one point, distinct from P , which does not have an infinite number of

congruent points distinct from P and in Q . This establishes the theorem.

THEOREM 3.—The set of limit points is transformed into itself by any transformation of the group.

The centers of the isometric circles consist of all points congruent to infinity. The transform of the center of an isometric circle is the center of another isometric circle or is ∞ itself.

Let P be a cluster point of the centers $g_1, g_2, \dots$. Then a transformation S which carries P into P' carries $g_1, g_2, \dots$ into $g_1', g_2', \dots$ with P' as cluster point. The points of the latter set, with the possible exception of one point at ∞ , are centers of isometric circles. Hence, P' is a limit point.

Furthermore, no point which is not a limit point is carried by S into a limit point, since otherwise S^{-1} would carry a limit point into a point not a limit point.

THEOREM 4.—If the set of limit points contains more than two points, it is a perfect set.

A set is perfect, by definition, if it has the following two properties: (1) each cluster point of the set belongs to the set; that is, the set is closed; and (2) each point of the set is a cluster point of points of the set; that is, the set is dense in itself.

That the set is closed follows at once. For, since each limit point contains an infinite number of centers of isometric circles in its neighborhood, a point at which limit points cluster has also an infinite number of centers of isometric circles in its neighborhood; hence, a cluster point of limit points is itself a limit point.

To establish the second property we must show that any limit point P has an infinite number of limit points in its vicinity. If P_1 and P_2 are two other limit points, at least one of them has an infinite number of transforms in the neighborhood of P (Theorem 2). As these transforms are limit points, the second property of the perfect set is established.

There are groups of transformations—the finite groups—with no limit points. Groups with a single limit point and groups with two limit points exist. A group other than these simple kinds has an infinite number of limit points. Furthermore, by a well-known property of perfect sets, the limit points are non-denumerable.

THEOREM 5.—If a closed set of points Σ , consisting of more than one point, is transformed into itself by all transformations of the group, then Σ contains all the limit points of the group.

Suppose, on the contrary, that there is a limit point P not belonging to Σ . Then, since Σ is closed, there is no point of Σ within a suitable neighborhood of P . Let P_1, P_2 be two points of Σ . At least one (Theorem 2) has transforms in the neighborhood of P . This contradicts the hypothesis that Σ is transformed into itself.

As an example of the use of the last theorem, suppose that all the transformations of the group are real. Then the real axis is always transformed into itself. It follows that the limit points all lie on the real axis.

19. Definition of the Region R .— R will consist of all that part of the plane which is exterior to the isometric circles of all the transformations of the group. More accurately, a point z will belong to R if a circle can be drawn with z as center which contains no point interior to an isometric circle. We thus rule out those limit points, if any, which are not themselves within or on an isometric circle but which have arcs of isometric circles in any neighborhood of them. Later, we shall adjoin to R a part of its boundary, but for the present it shall consist only of interior points. It may be a connected region, or it may comprise two or more disconnected parts. We see from (9) that it contains all of the plane lying outside a circle concentric with Q_0 and of radius 3ρ .

It is clear that no two points of R are congruent. A transformation T carries all points exterior to I_i into the interior of I_i' . Any transformation of the group, except the identical transformation, carries a point of R into an isometric circle and, hence, outside R ; so no point of R is congruent to another point of R .

20. The Regions Congruent to R .—If we apply to R the various transformations of the group, there results a set of congruent regions no two of which overlap (Theorem 1). Concerning the distribution of these regions, we have the following important theorem:

THEOREM 6.— R and the regions congruent to R form a set of regions which extend into the neighborhood of every point of the plane.

Suppose, on the contrary, that there is a point z_0 enclosed by a circle Q with z_0 as center and of radius r sufficiently small that Q contains neither points of R nor points congruent to points of R . Then, all transforms of Q contain neither points of R nor points congruent to points of R . In particular, Q and

its transforms contain the centers of no isometric circles, since these are congruent to ∞ , which is a point of R . The interior points of Q and of its transforms are ordinary points.

Since z_0 is not a point of R , z_0 is within or on the boundary of some isometric circle. Similarly, the center of each circle congruent to Q lies within or on the boundary of some isometric circle.

The proof consists in showing that there is a circle congruent to Q of arbitrarily large radius, which constitutes a contradiction. Let S be a transformation whose isometric circle I_i has z_0 for an interior or boundary point. The center of I_i is exterior to Q . Consider the circle Q_1 into which S carries Q . S is equivalent to an inversion in I_i , followed by a reflection in a line and possibly a rotation. The magnitude of Q_1 is determined by the inversion.

It is a matter of simple algebra to show that if the center of Q is on I_i , the radius of Q_1 is

$$r_1 = \frac{r}{1 - \frac{r^2}{r_i^2}},$$

where r_i is the radius of I_i . If z_0 lies within I_i , r_1 exceeds this value. Since $r_i < 2\rho$ (Equation (9)), and $r < r_i < 2\rho$, we have

$$r_1 > kr, \quad k = \frac{1}{1 - \frac{r^2}{4\rho^2}} > 1.$$

If we apply to Q_1 a transformation whose isometric circle has the center of Q_1 as an interior or boundary point, we get a circle Q_2 of radius r_2 where, since $r_1 > r$,

$$r_2 > \frac{r_1}{1 - \frac{r_1^2}{4\rho^2}} > kr_1 > k^2r.$$

Continuing in this manner, we prove the existence of a circle Q_n congruent to Q and of radius exceeding $k^n r$. By taking n large enough, Q_n will contain points of R exterior to the finite region in which the isometric circles lie. These points are congruent to points of Q , a contradiction which proves the theorem.

We can now establish the following result:

THEOREM 7.— R constitutes a fundamental region for the group.

We have already shown that no two points of R are congruent. We must show further that in any circle Q about a point P on the boundary of R there are points congruent to points of R . Let z_0 be a point of Q which lies in an isometric circle I . Then in the region common to Q and I , which contains no points of R , there are, by Theorem 6, points congruent to points of R . This establishes the theorem.

THEOREM 8.—Any closed region not containing limit points of the group is covered by a finite number of transforms of R (including possibly R itself). These regions fit together without lacunæ.

Let A be a closed region; for example, a region bounded by a simple closed curve, having no limit points of the group in its interior or on the boundary. Then there is a finite number of isometric circles containing points of A . For, if there is an infinite number, there are circles of arbitrarily small radius. Their centers then have a point of A as cluster point, contrary to hypothesis.

A transformation S carries R into a region R_s lying in I_s' the isometric circle of S^{-1} . If I_s' contains points of A , R_s may contain points of A ; if A is exterior to I_s' , then R contains no points of A . Hence, the number of regions congruent to R which lie wholly or in part in A is not greater than the number of isometric circles which contain points of A . This number is finite. Also, since there are points of R , or points congruent to points of R in the neighborhood of every point of A (Theorem 6), it follows that the regions fit together without lacunæ. A is completely covered, except, of course, for the boundaries separating the various regions.

THEOREM 9.—Within any region enclosing a limit point of the group, there lie an infinite number of transforms of the entire region R .

This theorem follows at once from the fact that there is an infinite number of isometric circles lying entirely within a given region enclosing a limit point. Each of these circles contains a region congruent to the entire region R , and the various transforms are different regions.

The preceding theorems furnish a picture of the transforms of R . R and the regions congruent to it fit together to fill up all that part of the plane which is composed of ordinary points. They cluster in infinite number about the limit points.

21 The Boundary of R .—A point on the boundary of R is a point P not belonging to R but such that in any circle with P as center there are points of R . P may be an ordinary point or a limit point. Obviously, P cannot lie within an isometric circle.

If P is an ordinary point, it lies on one or more isometric circles. Since there is but a finite number of isometric circles whose arcs lie in the neighborhood of an ordinary point, a circle Q can be drawn with P as center such that Q is exterior to all isometric circles other than those which pass through P .

In the most general case, a boundary point P belongs to one of the following three categories:

- (α) P is a limit point of the group;
- (β) P is an ordinary point and lies on a single isometric circle;
- (γ) P is an ordinary point and lies on two or more isometric circles. P is then called a "vertex."

It is desirable to include under (γ) the following special case: If P is the fixed point of an elliptic transformation of period two, so that, although P lies on a single isometric circle, it separates two congruent arcs on the circle, we shall classify P under (γ) rather than (β). The advantages of this classification will appear subsequently.

Concerning the boundary points of category (α), there is nothing to be added to the theorems on limit points already derived in Sec. 18. We shall show subsequently (Sec. 25) that groups exist for which the boundary points of R are all limit points. The groups of interest for our present theory, however, possess ordinary boundary points also.

(α) *The Sides.*—Consider a boundary point of category (β). Let P lie on I_1 , and let P' on I_1' be the point into which T carries P . We shall show that P' is also a boundary point of category (β). We put aside the case $P' = P$, a situation which can arise only if I_1 and I_1' coincide and P is a fixed point of the resulting elliptic transformation of period two; for this case has been included in (γ).

First, P' is within no isometric circle. Suppose P' to be within I_s ; then S magnifies lengths in the neighborhood of P' . But, since T carries P into P' without alteration of lengths, ST magnifies lengths in the neighborhood of P . Then P is within I_n , which is contrary to the hypothesis that P is a boundary point.

Second, P' does not lie on an isometric circle other than I'_i . For, if P' lies on I_s , the transformation ST effects no alteration in lengths at P . Then P is on I_{st} , which is contrary to the hypothesis that P lies on a single isometric circle. It follows from these facts that P' is a boundary point of category (β).

There is no isometric circle in the neighborhood of P other than I_i . It is clear, then, that the points on I_i in the neighborhood of P are likewise boundary points of category (β); so, consequently, are the congruent points on I'_i . We thus have as a part of the boundary an arc of I_i and the congruent arc on I'_i . These arcs may consist of the entire circles or they may terminate in points of category (α) or (γ). Since the arcs lie on isometric circles, they are of equal length. We have, then, the following theorem:

THEOREM 10.—*The boundary points of R of category (β) form a set of bounding circular arcs, or sides, which are congruent in pairs. Two such congruent sides are equal in length.*

(b) *The Vertices.*—There remain for consideration the boundary points of category (γ). Through a point P of this category there pass a finite number of isometric circles. Let Q be a circle about P sufficiently small that all isometric circles other than those through P are without Q and such that any points of intersection of the circles through P , other than P itself, lie without Q . The isometric circles through P divide Q into a finite number of parts. One of these parts A , owing to the assumption that P is a boundary point, belongs to R . The two arcs which bound A , on I_i and I_s , say, are a part of the boundary of R . The points of these arcs other than P belong to category (β). That is, at a vertex two sides of R meet.

Now make the transformation T , P being carried into P' on I'_i . By reasoning almost identical with that employed in the preceding case, we show that P' is a vertex. We can show, in fact, that P and P' lie on the same number of isometric circles. Let P lie on the isometric circles of T , S , U , . . . ; then P' lies on the isometric circle of T^{-1} , and also on those of ST^{-1} , UT^{-1} , . . . , as we see on considering the way in which lengths in the neighborhood of P' are affected by the transformations. These transformations are different—thus, $ST^{-1} = UT^{-1}$ implies $S = U$ —hence, their isometric circles are different. Hence, as many isometric circles pass through P' as through P . Interchanging the rôles of P and P' , as many pass through P as through P' .

(c) *Extension of R .*—We shall find it convenient to add to the region R certain points of its boundary. Of two congruent sides, one, exclusive of the end points, may be added without including points congruent to points previously in R . A vertex where bounding arcs meet may be congruent to several other vertices, one of which may be adjoined. The resulting region is still a fundamental region.

22. Example.—*A Finite Group.*—The following example of a finite group illustrates parts of the preceding discussion. The transformations

$$T_0 = z, \quad T_1 = \frac{2z-1}{3z-2}, \quad T_2 = \frac{z}{5z-1}, \\ T_3 = \frac{3z-1}{7z-2}, \quad T_4 = \frac{2z-1}{7z-3}, \quad T_5 = \frac{3z-1}{8z-3},$$

constitute a group. This is, in fact, the group got by transforming the group of anharmonic ratios (Sec. 13 (2)) by $G = 1/(z+2)$ so that ∞ shall not be a fixed point of any transformation. The isometric circles are

$$I_1: |z - \frac{3}{8}| = \frac{1}{8}, \quad I_2: |z - \frac{1}{2}| = \frac{1}{4}, \quad I_3: |z - \frac{1}{4}| = \frac{1}{4}, \quad I_4: |z - \frac{3}{4}| = \frac{1}{4}, \quad I_5: |z - \frac{3}{8}| = \frac{1}{8}.$$

These are shown in Fig. 15.

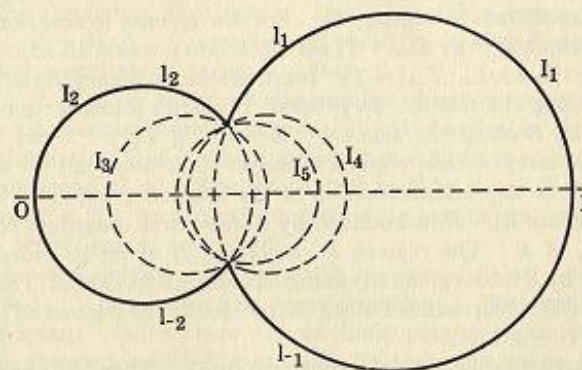


FIG. 15.

The fundamental region R is bounded by arcs of I_1 and I_2 . T_1 is an elliptic transformation of period two [$a+d=0$], one of the fixed points being 1. T carries the upper half I_1 of the bounding arc on I_1 into the lower half I_{-1} . Similarly, T_2 is of period two, the upper and lower sides, I_2 and I_{-2} being congruent. There are thus two pairs of equal congruent sides.

Two congruent vertices are $\frac{5}{14} \pm i\sqrt{\frac{3}{14}}$, through each of which pass all five isometric circles. The points 0 and 1 are also vertices.

R and the five regions congruent to R fill up the whole plane without overlapping. The congruent regions are not drawn, but it is not difficult

to verify that the six regions consist of the six parts into which the plane is divided by the complete isometric circles I_1 , I_2 , and I_3 .

23. Generating Transformations.—In Sec. 13, mention was made of the group resulting from combining a finite or infinite number of linear transformations in all possible ways. We shall now raise the question whether a knowledge of the fundamental region R furnishes any information as to a set of generating transformations of the group.

Let $l_1, l_{-1}, l_2, l_{-2}, \dots$ be the congruent pairs of sides of R . Let T_1 carry l_1 into l_{-1} ; T_2 carry l_2 into l_{-2} ; and so on. We shall show that under certain circumstances $T_1, T_2, \dots$ form a set of generating transformations for the group.

It is clear that the group formed by all possible combinations of $T_1, T_2, \dots$, and their inverses contains only transformations of the original group. That is, the group generated by $T_1, T_2, \dots$, which we shall designate by Γ_i , is a subgroup of the original group. The question at issue is whether Γ_i contains all the transformations of the group or only a part of them.

Let us consider the way in which R is transformed by the transformations of the group Γ_i . For convenience in notation, let us represent T_n^{-1} by T_{-n} . T_1 carries R into a region R_1 adjoining R along the side l_{-1} ; $T_{-1} (= T_1^{-1})$ carries R into a region R_{-1} adjoining R along the side l_1 . In general, T_n carries R into a region R_n adjoining R along the side l_{-n} . Hence, $T_1, T_2, \dots$ and their inverses carry R into regions adjacent to R along all its sides.

Let S be any transformation of the original group, and let S carry R into R_s . R_s is bounded by circular arcs congruent to the sides l_n of R . The regions R_n adjoining R along its sides are carried by S into regions adjoining R_s along all its sides. That is, $R_s (= S(R))$ is surrounded along all its sides by the regions $ST_n(R)$, $n = \pm 1, \pm 2, \dots$

In particular, let S belong to the group Γ_i . Then, ST_n belongs to Γ_i , and the regions surrounding R_s are transforms of R by transformations of the group Γ_i . The transforms of R by the group Γ_i are entirely surrounded by other such transforms, the regions fitting together, without lacunae and with no free sides, to form one or more networks of regions. If these networks fill up the whole plane, Γ_i coincides with the original group; otherwise it does not.

Let A be a closed region containing no limit points of the group and having points in common with the region R . Then R and a

finite number of its transforms by the group Γ_i cover A completely. For, along the sides of R which lie in A , we can adjoin regions of the network; along the sides of the new regions which lie in A , we can adjoin other regions of the network; and so on. We can continue this process as long as there are any free sides lying in A . But A is covered completely by a finite number of regions congruent to R (Theorem 8); hence, this process must end in a finite number of steps.

By similar reasoning, if the region A contains points of any region R_i congruent to R by a transformation of Γ_i , A can be covered by a finite number of regions congruent to R by the transformations of Γ_i .

We can now prove the following theorem:

THEOREM 11.—*If some point of R can be joined to all its congruent points by curves not passing through limit points, then the transformations $T_1, T_2, \dots$ by which the sides of R are congruent constitute a set of generating transformations for the group.*

Let S be any transformation of the group. Let z_0 be a point of R which can be joined to any congruent point by a curve not passing through a limit point. Let C be such a curve joining z_0 to $z_0' = S(z_0)$. C can be embedded in a closed region A consisting entirely of ordinary points.

A is covered by R and a finite number of regions congruent to R by transformations of the group Γ_i . In particular, there is a transformation T of Γ_i which carries R into a region covering the neighborhood of z_0' . The transforms of R by S and T overlap, whence $S = T$. Any transformation of the original group belongs to Γ_i , which was to be proved.

There are certain cases in which we can state at once that $T_1, T_2, \dots$ are generating transformations. The first is the finite group. Since there are no limit points, the conditions of the theorem hold. For example, in the group given in Sec. 22, T_1 and T_2 are generating transformations. The reader can verify that their combinations give the remaining transformations.

Again, the distribution of the limit points may be such that any two points which are not limit points may be joined by a curve not passing through the limit points. Then, obviously, the conditions of the theorem hold. The simplest examples are the groups with a finite number (one or two) of limit points.

24. Cyclic Groups. **DEFINITION.**—*A cyclic group is a group generated by a single transformation.*

If T is the transformation which generates the group, then the group consists of the transformations $\dots T^{-2}, T^{-1}, 1, T, T^2, \dots$. The groups whose fundamental regions are shown in Figs. 11, 12, 14 are cyclic groups. In the examples, all the transformations have a fixed point at infinity. In the present section, we shall examine the cyclic groups arising when T is one or another of the various types of transformations and where the fixed points are finite. A knowledge of cyclic groups is important because of the fact that every group contains cyclic subgroups. For, the group generated by any transformation of the original group belongs to that group.

If there are two fixed points, ξ_1 and ξ_2 , T can be written in the form (Sec. 6, Equation 24).

$$\frac{z' - \xi_1}{z' - \xi_2} = K \frac{z - \xi_1}{z - \xi_2}, \quad K \neq 1, \quad (11)$$

and the general transformation T^n is

$$\frac{z' - \xi_1}{z' - \xi_2} = K^n \frac{z - \xi_1}{z - \xi_2}. \quad (12)$$

On solving for z' , we have

$$z' = T^n(z) = \frac{(K^n \xi_2 - \xi_1)z + (1 - K^n)\xi_1 \xi_2}{(K^n - 1)z + \xi_2 - K^n \xi_1}.$$

We find the determinant to be $K^n(\xi_1 - \xi_2)^2$, so we must divide numerator and denominator by $K^{n/2}(\xi_1 - \xi_2)$ to render the determinant 1. The isometric circle I_n of T^n is

$$\left| z + \frac{\xi_2 - K^n \xi_1}{K^n - 1} \right| = \left| \frac{K^{n/2}(\xi_1 - \xi_2)}{K^n - 1} \right| = \left| \frac{\xi_1 - \xi_2}{K^{n/2} - K^{-n/2}} \right|. \quad (13)$$

(a) *Hyperbolic and Loxodromic Cyclic Groups.*—If T is hyperbolic, K is real and $|K| \neq 1$; then the multiplier K^n of T^n , $n \neq 0$, is likewise real and in absolute value unequal to 1; whence all transformations of the group are hyperbolic. If K is not real and $|K| \neq 1$, then $|K^n| \neq 1$, $n \neq 0$, although in certain cases K^n may be real; hence, if T is loxodromic all the transformations of the group are loxodromic or hyperbolic.

We found in Sec. 11 (b) that a fixed point cannot lie on either of the isometric circles I_n, I_n' of a hyperbolic or loxodromic transformation T^n and its inverse. Clearly, there can be no fixed points outside both I_n and I_n' , for T^n shifts the position of

such a point. Each I_n contains a fixed point. As n increases, the radius of I_n approaches zero; so each fixed point is a limit point of the group. An application of Theorem 5, where we let the set Σ consist of the two fixed points, shows that there are no further limit points.

In studying the arrangement of the isometric circles we shall make use of the following proposition:

THEOREM 12.—Let I_t, I_s, I_t', I_s' be the isometric circles of $T, S, S^{-1}, U = ST$, respectively. If I_t and I_s' are exterior to one another, then I_u is contained in I_s .

If I_t and I_s' are tangent externally, then I_u lies in I_s and is tangent internally.

The proof is simple. Suppose the circles not tangent; and let z be a point outside (or on) I_s . Then S carries z into a point z' within (or on) I_s' with decrease of lengths (or without alteration of lengths). Now z' is outside I_t ; so T transforms z' with decrease of lengths. Hence, U transforms z with decrease of lengths, whence z is outside I_u . Since every point on or outside I_s is also outside I_u , the latter circle is contained in the former.

If I_t and I_s' are tangent externally at a , the preceding reasoning holds except at the point a_0 on I_s which S carries into a . TS makes no alteration of lengths at a_0 , whence a_0 lies on I_u .

If, in the cyclic group, T is hyperbolic or if T is loxodromic and $|a + d| \geq 2$, I and I' , the isometric circles of T and T^{-1} , are exterior to one another (possibly tangent). We now show that I encloses I_2 , I_2 encloses I_3 , etc., and, likewise, I' encloses I_2', I_2' encloses I_3' , and so on. We see at once from Theorem 12, taking $S = T$, that I_2 is contained in I . Similarly, from the product $T^{-1}T^{-1}$, I_2' lies in I' .

We establish the general relation by induction. Suppose that the circles are arranged as stated up to I_n and I_n' , and consider I_{n+1} . We write $T^{n+1} = TT^n$. By hypothesis I_n' , the isometric circle of T^{-n} , lies in I' , and hence is exterior to I . It follows from Theorem 12 that I_{n+1} lies within I_n . By identical reasoning I_{n+1}' is contained in I_n' .

It follows from the preceding that I and I' contain all other isometric circles, whence the fundamental region R is the region lying outside these two circles. In Fig. 16, the isometric circles are drawn for a group generated by a hyperbolic transformation with $K = 4$.

The situation is somewhat different for a loxodromic transformation for which $|a + d| < 2$. In Fig. 17 the isometric circles are shown for the group generated by $T = 1/(z + 1)$. Here $a + d = i$, whence T is loxodromic. The congruent boundaries of R are connected by the transformations T and T^2 .

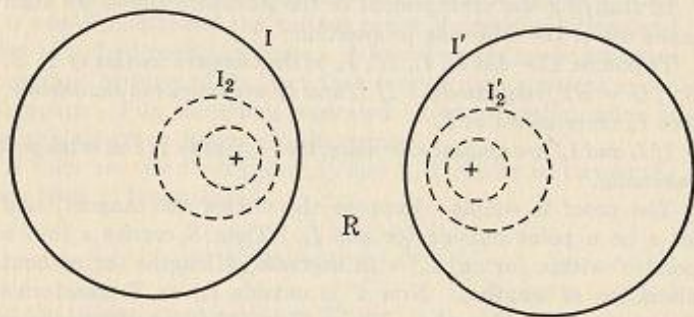


FIG. 16.

According to Theorem 11, these are generating transformations for the group. This example illustrates the fact that the generating transformations found by an application of Theorem 11 are not necessarily the best obtainable, that some of the transformations found in this way may be consequences of others.

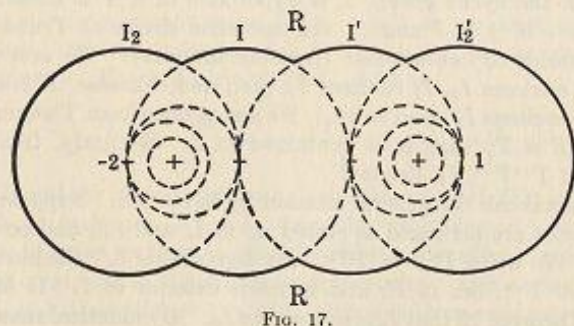


FIG. 17.

A simpler fundamental region for this case, although not bounded by isometric circles, is got as follows. The loxodromic transformation may be written $T = UV$, where V and U are hyperbolic and elliptic transformations, respectively, with the same fixed points as T . Let I and I' be the isometric circles of V and V^{-1} . Then, it is readily shown that the part of the plane exterior to I and I' is a fundamental region for the group generated by T .

(b) *Elliptic Cyclic Groups*.—If T is elliptic, all isometric circles pass through the fixed points. Here $K = e^{i\theta}$, and unless θ is commensurable with π , the group is continuous. In Fig. 18, the isometric circles are drawn for $K = e^{6\pi i/5}$. The group may be generated, as the figure shows, by T^2 which has the multiplier $e^{2\pi i/5}$.

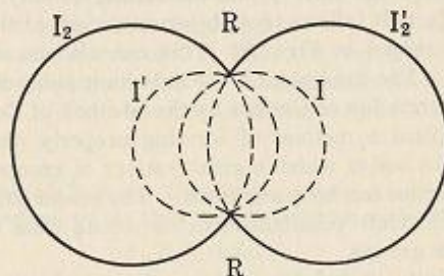


FIG. 18.

(c) *Parabolic Cyclic Groups*.—If T is parabolic, the transformation may be written (Sec. 10, Equation 37)

$$\frac{1}{z' - \xi} = \frac{1}{z - \xi} + c, \quad c \neq 0. \quad (14)$$

Then T^n is the transformation

$$\frac{1}{z' - \xi} = \frac{1}{z - \xi} + nc, \quad (15)$$

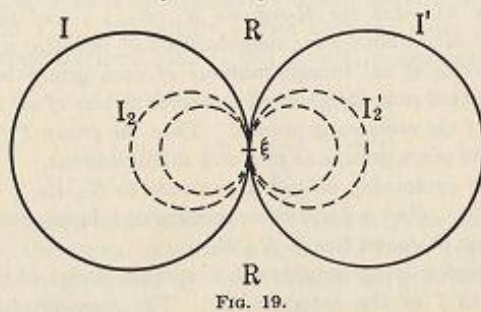


FIG. 19.

or

$$z' = \frac{(nc\xi + 1)z - nc\xi^2}{ncz + 1 - nc\xi}. \quad (16)$$

The determinant in (16) is 1 and the isometric circle I_n is

$$\left| z - \left(\xi - \frac{1}{nc} \right) \right| = \frac{1}{|nc|}. \quad (17)$$

We see from (17) that ξ lies on I_n , and that the center of I_n , namely $\xi - \frac{1}{nc}$, lies on the line joining ξ to $1/c$. Hence, all isometric circles have a common tangent at ξ . If n is positive, the center of I_n lies on one side of ξ ; if negative, on the other. As n increases in absolute value, the radius steadily diminishes. From these facts it follows that the arrangement of the isometric circles is as shown in Fig. 19. This can also be shown from Theorem 12. The fixed point is the only limit point of the group.

25. The Formation of Groups by the Method of Combination. We here explain a method of forming properly discontinuous groups, by the use of which a great variety of groups with very diverse properties can be constructed. The reader will gain some idea of the intricate possibilities in the broad class of properly discontinuous groups.

Given a finite or infinite number of properly discontinuous groups $\Gamma_1, \Gamma_2, \dots$. Let the transformations of these groups be combined in all possible ways to form a group Γ . We get two kinds of transformations: (a) those belonging to the original groups; and (b) *cross-products* resulting from the combination of transformations not all belonging to the same original group. The resulting group may be continuous or discontinuous. In certain cases, however, we are able to state that the group is properly discontinuous and to specify its fundamental region R .

THEOREM 13.—Let the R -regions $R_1, R_2, \dots$ of the groups $\Gamma_1, \Gamma_2, \dots$ all contain some neighborhood of infinity; and let the isometric circles of all transformations of each group be exterior (possibly tangent externally) to all isometric circles of all the transformations of the remaining groups. Then the group Γ formed by combining the given groups is properly discontinuous.

The region containing all points common to $R_1, R_2, \dots$ is the region R for Γ . Here a point is not counted as belonging to R unless a region about the point lies in $R_1, R_2, \dots$.

R is the region lying outside the isometric circles of the transformations of Γ of the category (a). The cross-products have not been taken into account. We now show that the isometric circles of the transformations (b) contain no points of R .

Consider the general cross-product

$$U = S_n S_{n-1} \cdots S_2 S_1.$$

Here we suppose that the two successive transformations S_n, S_{n+1} do not belong to the same one of the groups $\Gamma_1, \Gamma_2, \dots$;

otherwise, we combine them into one transformation. We apply the transformation U to a point z_0 of R . Let I_i, I_i' represent the isometric circles of S_i, S_i^{-1} . Since z_0 is exterior to I_1 , S_1 carries z_0 into z_1 within I_1' . Since I_1' is exterior to I_2 , z_1 is exterior to I_2 and S_2 carries z_1 into z_2 within I_2' ; and so on. At each step, lengths in the neighborhood of the point are decreased. So U decreases lengths in the neighborhood of z_0 ; hence z_0 is exterior to the isometric circle of U . Thus the region R of the theorem is the region lying outside all the isometric circles belonging to the group Γ . The existence of R shows that Γ is properly discontinuous.

By the use of Theorem 12, employing reasoning similar to that of the preceding section, we can show that the isometric circles of

$$S_1, S_2 S_1, S_3 S_2 S_1, \dots, U$$

form a sequence such that each circle encloses the circle which follows it. It follows that the isometric circle of a cross-product lies within one of the isometric circles of the original groups, and hence has no bearing on the construction of R .

A few examples wherein we actually construct combination groups will throw some light on the various forms which the region R can take.

Given two equal circles, we can set up infinitely many transformations T such that the given circles are the isometric circles of T and T^{-1} . It is easily shown that the most general linear transformation such that I_i and I_i' are, respectively,

$$|z - q| = r, \quad |z - q'| = r,$$

is

$$T = \frac{q'z - (qq' + r^2 e^{i\theta})}{z - q}, \quad (18)$$

where θ is any real quantity. If I_i and I_i' are exterior to one another, the region exterior to I_i and I_i' is the region R for the cyclic group generated by T . We shall use groups of this sort in applying Theorem 13.

(a) Given $2n$ circles $I_1, I_1'; \dots; I_n, I_n'$, which are equal in pairs, and which are exterior to one another or are externally tangent. We set up the transformation T_i so that T_i and T_i^{-1} have the isometric circles I_i and I_i' . Then the group generated by $T_1, T_2, \dots, T_n$ has for a fundamental region the region outside the $2n$ circles.

If no circles are tangent, R is a connected region, although not simply connected. Its boundary consists entirely of sides, there being no boundary points of categories (α) or (γ) (Sec. 21).

By making the circles tangent in various ways, the region exterior to the circles can be separated into several regions. Then R will consist of several parts.

(b) In the same way, we can select an infinite number of circles equal in pairs and construct a group. Since we are assuming that the isometric circles lie in some finite domain, we shall have circles of arbitrarily small radius.

These circles can be put together in such a way that R is composed of an infinite number of separate parts.

(c) That R can be a region whose boundary points are all limit points is shown by the following example:

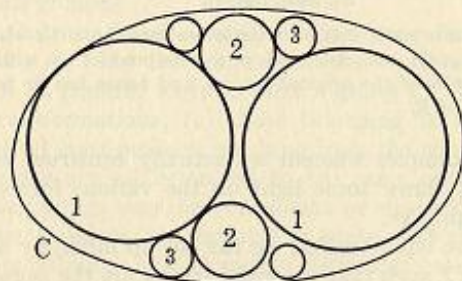


FIG. 20.

Let C be a closed curve (Fig. 20). Without going into detail, it is seen that pairs of isometric circles with which to form a combination group can be put into the interior of C in such a manner that every point within or on C which is not within a circle has an infinite number of circles in its neighborhood. Then R comprises the exterior of C . The boundary of R , namely C itself, is composed entirely of limit points.

Constructions of this sort for other and more complicated forms of the region R will occur to the reader.

An Extension. Schottky Groups.—Results analogous to Theorem 13 can be derived for other kinds of fundamental regions. Let $\Gamma_1, \dots, \Gamma_m$ be groups with fundamental regions $F_1, \dots, F_m$. Let F_i contain in its interior all points of the plane not interior to $F_j (i \neq j)$. If the combination group Γ be formed we can show that the region F consisting of all points

common to $F_1, \dots, F_m$ is a fundamental region for Γ . The boundary of F consists of the boundaries of $F_1, \dots, F_m$.

In the neighborhood of a point P on the boundary of F_i there exist points not belonging to F_i which are congruent to points of F_i . Further, these points can be so chosen that the congruent points in F_i lie as near the boundary of F_i as we wish and so belong to F . The second requirement of the fundamental region is satisfied.

Obviously, no two points of F are congruent by transformations of the original group. We now consider the effect of a cross-product on a point z_0 of F . Let $U = S_n \dots S_1$, where $S_i (\neq 1)$ belongs to the group Γ_{n_i} and Γ_{n_i} and $\Gamma_{n_{i+1}}$ are different groups. S_1 carries z_0 into a point z_1 outside F_{n_1} and hence in F_{n_2} ; S_2 carries z_1 into a point z_2 outside F_{n_2} and hence in F_{n_3} ; etc. Finally, S_n carries z_{n-1} into a point z_n outside F_{n_n} ; that is, U carries z_0 into a point outside F , and the first property of the fundamental region is established.

The *Schottky group*¹ is constructed as follows: Let $Q_1, Q_1'; \dots; Q_m, Q_m'$ be m pairs of circles external to one another. Let T_i be a linear transformation (loxodromic or hyperbolic) carrying Q_i into Q_i' in such a way that the exterior of Q_i is carried into the interior of Q_i' . T_i generates a cyclic group Γ_i for which all that part F_i of the plane exterior to Q_i and Q_i' is a fundamental region. Here F_i contains all that part of the plane not contained in $F_j (j \neq i)$. The Schottky group Γ is got by combining $\Gamma_1, \dots, \Gamma_m$. It has as fundamental region all that part of the plane exterior to the $2m$ circles. Γ is generated by the transformations $T_1, \dots, T_m$.

In a subsequent chapter (Chap. X) there will arise groups generated as is the Schottky group except that $Q_1, \dots, Q_m'$ are closed curves which are not necessarily circles. Such a combination group is called a "*group of Schottky Type*."

26. Ordinary Cycles.—Returning now to the fundamental region R for the general properly discontinuous group, we shall make a study of the vertices. Let A_1 be a vertex. If either of the sides which meet in A_1 is carried into its congruent side, A_1 is carried into a vertex at the extremity of this latter arc. These congruent vertices may be carried into others, with the result that A_1 may be congruent to several of the vertices of R .

¹ *Crelle's Jour.*, Vol. 101, pp. 227-272, 1887.

We may find the vertices congruent to A_1 by the following method: Let us think of the boundary of R as being traced in a positive sense; that is, with the region on the left. In passing through a vertex, we proceed along one side to the vertex and then proceed along a second side from the vertex. We consider the vertex as the end of the former side and the beginning of the latter. When a side is carried into its congruent side, the beginning and end of the former are carried into the end and beginning, respectively, of the latter; this results from the fact that the direction around the isometric circle is reversed (Sec. 11 (a)).

Let l_1 be the side beginning at A_1 (Fig. 21). This side is carried by a transformation T_1 into the congruent side l_{-1} , A_1 being carried into A_2 at the end of l_{-1} . There is a side l_2 beginning at A_2 . The arc l_2 is carried by some transformation T_2 into the congruent side l_{-2} , A_2 being carried into A_3 at the end of l_{-2} . We can continue in this manner, getting other congruent vertices until we return to A_1 and the side l_1 .

We will return to A_1 in a finite number of steps. Suppose, on the contrary, that an infinite number of vertices $A_2, A_3, \dots$ are congruent to A_1 . The transformation S_n which carries A_1 to A_n has an isometric circle passing through A_1 ; for if A_1 is outside the isometric circle of S_n , A_n is within the isometric circle of S_n^{-1} , which is impossible. Then the isometric circles of $S_2, S_3, \dots$, an infinite number, pass through A_1 , which is contrary to the hypothesis that A_1 is an ordinary point. Hence, in applying the process just explained, we encounter a finite number of vertices $A_2, A_3, \dots, A_m$, congruent to A_1 and then return to A_1 .¹

DEFINITION.—A complete set of congruent vertices of a fundamental region is called an ordinary cycle.

We shall show presently that there are no vertices of R congruent to A_1 other than those just found; whence $A_1, A_2, \dots, A_m$ constitute a cycle.

Let $T_1, T_2, \dots, T_m$ be the transformations in the preceding treatment by which we carry A_1 to A_2, A_2 to $A_3, \dots, A_m$ to A_1 , respectively. Some of these transformations, it will be noted, may be inverses of the others, but each connects a pair of bound-

¹ We continue until we encounter the vertex A_1 followed by the side l_1 . We may encounter A_1 once before this happens, in the special case that the sides at A_1 are tangent. See Fig. 24 p. 73.

ing arcs. The transformation $S = T_m \cdots T_2 T_1$ carries A_1 into itself. It may happen that S is the identical transformation. If not, S is an elliptic transformation. For, A_1 is a fixed point of S ; and the fixed points of hyperbolic or loxodromic transformations lie within isometric circles and the fixed point of a parabolic transformation (Fig. 19) is a limit point of the group.

Consider now the way in which the transforms of R fit together at A_1 . The transformations $T_m, T_m T_{m-1}, \dots, T_m \cdots T_2, T_m \cdots T_1 (= S)$ carry $A_m, A_{m-1}, \dots, A_2, A_1$, respectively, into A_1 . Further, the regions $R_m, R_{m-1}, \dots, R_2, R_1$, respectively, into which these transformations carry R fit together

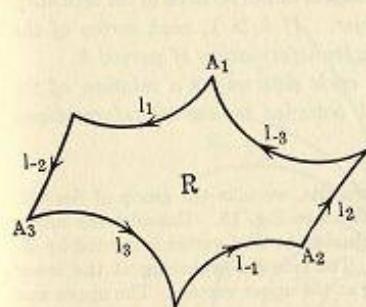


FIG. 21.

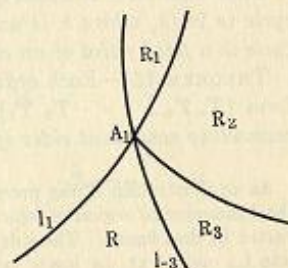


FIG. 22.

at A_1 (Fig. 22). Thus, T_m carries A_m into A_1 , R_m being adjacent to R along the arc l_{-m} ending at A_1 . In general, T_i carries A_i into A_{i+1} and carries R into a region abutting on R along a side ending in A_{i+1} , whence the transforms of these two regions by $T_m \cdots T_{i+1}$, namely R_i and R_{i+1} , are adjacent along an arc issuing from A_1 . In proceeding counterclockwise around A_1 , starting from R , we encounter in order the adjacent regions $R_m, R_{m-1}, \dots, R_2, R_1$. The curvilinear angle of R at A_n is carried into an equal angle of R_n at A_1 .

Since there can be no overlapping of congruent regions, there are two possibilities. First, R_1 may coincide with R and the regions $R, R_m, \dots, R_2$ completely fill up the angle about A_1 . Then, S is the identical transformation. The sum of the angles at the vertices $A_1, A_2, \dots, A_m$ is equal to 2π .

If R_1 does not coincide with R , S is an elliptic transformation. Now, an elliptic transformation with multiplier $K = e^{i\theta}$ amounts to a rotation in the neighborhood of the fixed point through the

angle $\theta \neq 0$. Carrying R into R_1 requires (Fig. 22) that θ be equal to the sum of the angles at the vertices. On applying S , R , $R_m, \dots, R_2$ are carried into adjacent regions filling out more of the region about A_1 counterclockwise from R_1 . After a finite number, k , of applications of S , the region about A_1 is completely filled up and $S^k(R)$ coincides with R .

It is now clear that there are no vertices of R congruent to A_1 , other than $A_2, \dots, A_m$. For, a transformation carrying any other vertex A to A_1 carries R into a region overlapping the regions which fill out the angle about A_1 , which is impossible.

We state the preceding results in the following form:

THEOREM 14.—*The sum of the angles at the vertices of an ordinary cycle is $2\pi/k$, where k is an integer. If $k > 1$, each vertex of the cycle is a fixed point of an elliptic transformation of period k .*

THEOREM 15.—*Each ordinary cycle determines a relation of the form $(T_m T_{m-1} \dots T_2 T_1)^k = 1$ satisfied by the transformations connecting congruent sides of R .*

As an illustration of the preceding results, we take the group of Sec. 22, the fundamental region of which is shown in Fig. 15. Consider the upper vertex in that figure. The side l_1 beginning at that vertex is carried by T_1 into l_2 ending at the lower vertex. The side l_2 beginning at the lower vertex is carried by T_2^{-1} into l_1 ending at the upper vertex. The upper and lower vertices thus constitute a cycle.

The transformation $S = T_2^{-1}T_1$ is clearly not the identical transformation, since the sum of the angles at the two vertices is less than 2π ; hence, S is an elliptic transformation with the upper vertex as fixed point. We find readily that $S = T_4$, and that $S^2 = 1$. The sum of the angles at the two vertices is $2\pi/3$.

Again the origin is a vertex. The arc l_2 beginning at the origin is carried by T_2 into the arc l_1 ending at the origin. Hence, the origin alone constitutes a cycle. The angle there is π ; whence $T_2^2 = 1$.

Similarly, the vertex at the point 1 constitutes a cycle. There are thus three cycles.

27. Parabolic Cycles.—If a side of R terminates in a limit point P_1 various situations may arise. It may happen that there is no other side terminating in P_1 . Let us suppose that two sides meet in P_1 ; and let us apply the method given in the preceding section for getting points congruent to P_1 .

Let l_1 be the side beginning at P_1 . Then l_1 is carried by a transformation T_1 into a side l_2 ending at a limit point P_2 . Let l_2 be the side, if any, beginning at P_2 , and let T_2 carry l_2 into the side l_3 ending at P_3 ; and so on.

In the application of this process there are three situations that may arise: (1) the process may be terminated at some stage by arriving at a point at which no side begins; (2) the process may continue ad infinitum without a return to P_1 ; (3) after arriving at a finite number of congruent points $P_2, P_3, \dots, P_m$, we may return to P_1 and the side l_1 . It is not difficult to set up combination groups exemplifying each possibility.

If (3) holds, we say that $P_1, P_2, \dots, P_m$ constitute a *parabolic cycle*, and each point of the cycle is called a "*parabolic point*."

In many of the groups to be studied subsequently R has a finite number of sides and the only limit points on the boundary

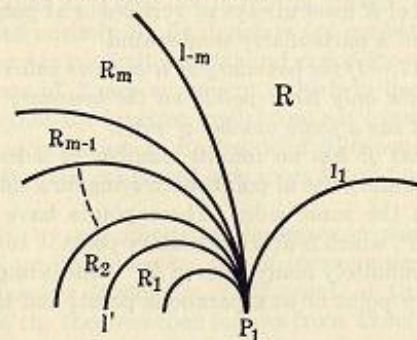


FIG. 23.

are points where two sides meet. These conditions rule out cases (1) and (2); whence the limit points on the boundary are parabolic points arranged in cycles.

The transformation $S = T_m T_{m-1} \dots T_2 T_1$ carries P_1 into itself; whence S is either elliptic or parabolic. The reasoning of the preceding section can be repeated word for word to show that $T_m, T_m T_{m-1}, \dots, T_m \dots T_2, T_m \dots T_1$ carry $P_m, P_{m-1}, \dots, P_2, P_1$, respectively, into P_1 and carry R into regions $R_m, R_{m-1}, \dots, R_2, R_1$ fitting together along arcs issuing from P_1 in the order shown in Fig. 23. The angle between the sides of R_i which meet at P_1 is equal to the angle between the sides of R which meet at P_1 .

S carries R into R_1 , the side l_1 being carried into the side l_1' which separates R_1 and R_2 . If S is elliptic l_1 and l_1' meet at an angle different from zero. By repeating S a finite number of times, the neighborhood of P_1 is covered by a finite number of regions, which is contrary to Theorem 9. Hence S is parabolic.

Since S is parabolic, l_1 and l' are tangent at the fixed point P_1 . Then, the arcs bounding the intervening regions are also tangent to l_1 . The angle between the sides which meet at each point of the cycle is zero.

By repeated application of S to the regions in Fig. 23, we get infinitely many other regions with two sides meeting at P_1 , the sides being tangent to l_1 .

THEOREM 16.—*The sides of R which meet at a parabolic point are tangent. There is an infinite number of regions congruent to R each having two sides which meet at the parabolic point and are tangent to the sides of R .*

If the sides of R meet always at vertices or at parabolic points, the region is of a particularly simple kind.

THEOREM 17.—*If the boundary of R consists entirely of ordinary points, or if the only limit points on the boundary are parabolic points, then R has a finite number of sides.*

Suppose that R has an infinite number of sides. Let $z_1, z_2, \dots$ be an infinite suite of points each lying on a side of R and no two lying on the same side. These points have at least one cluster point P , which is also a boundary point. In the neighborhood of P lie infinitely many sides of R . This is impossible either at an ordinary point or at a parabolic point; and the theorem is established.

28. Function Groups.—The ordinary points of a properly discontinuous group either form a single connected region, or two-dimensional continuum, Σ , or else are separated by the limit points into two or more two-dimensional continua $\Sigma, \Sigma_1, \Sigma_2, \dots$

DEFINITION.—*A properly discontinuous group will be called a function group if one of the connected regions Σ into which the limit points separate the plane is carried into itself by all the transformations of the group.*

In a finite group or a group with one or two limit points the ordinary points constitute a single connected region which is carried into itself by all transformations, whence the group is a function group. A group of the sort shown in Fig. 20 is not a function group. It is the function groups only that will play a part in our subsequent theory.

THEOREM 18.—*In the region Σ of a function group lies a part R_0 of the region R , and R_0 is a fundamental region.*

Not all of R can be exterior to Σ , for then the transforms of R would all be exterior to Σ , contrary to Theorem 6. Let P be

a boundary point of R_0 . If P is a limit point, there are points congruent to points of R_0 in the neighborhood of P by Theorem 2. If P is an ordinary point on the boundary of R_0 , it lies in Σ and there are points congruent to points of R in its neighborhood. But these points are congruent to points of R_0 since the transforms of all other points of R are exterior to Σ . Since obviously no points of R_0 are congruent, R_0 is a fundamental region for the group.

THEOREM 19.—*The sides of R_0 are congruent in pairs; and the transformations connecting congruent sides of R_0 form a set of generating transformations for the function group.*

It is clear that R_0 is bounded in part by sides. For, otherwise, R_0 is bounded entirely by limit points (or consists of the whole plane if there are no limit points) and coincides with Σ . Then, no two points of Σ are congruent, which is impossible. (We assume here that the function group does not consist solely of the identical transformation.) Each side of R_0 is congruent to some side of R . But the congruent side must lie in Σ and so is a side of R_0 .

Since Σ is, by hypothesis, a connected region, any interior point of R_0 can be joined to any of its congruent points by a curve not passing through a limit point of the group. The latter part of the theorem then follows from Theorem 11.

THEOREM 20.—*If a function group possesses a fundamental region F which, together with its boundary, consists of interior points of Σ , and whose transforms cover the neighborhood of each of its boundary points, then the boundary of R_0 consists of interior points of Σ , and R_0 has a finite number of sides.*

A finite number of transforms of R_0 cover F and its boundary completely. Let us carry the portion of F in each region R_i into R_0 by means of the transformation which carries R_i into R_0 . These transforms of parts of F do not overlap, since no two points of F are congruent.

Furthermore, they fill R_0 completely without lacunæ. Suppose, on the contrary, that an interior point z_0 of R_0 is exterior to all the transforms of parts of F . Let z_1 be the nearest point of one of the parts in R_0 . Then in any neighborhood of z_1 are points which are not covered and which are, therefore, not congruent to points of F . On carrying the particular part back to F , z_1 goes into a boundary point of F which has in its neighborhood points not congruent to points of F . This is impossible.

If R_0 has a limit point on its boundary, so has one of the parts covering R_0 and there is a congruent limit point on the boundary of F , contrary to hypothesis.

We may now imbed the region R_0 in a closed region A consisting entirely of ordinary points. Only a finite number of isometric circles enclose points of A . As isometric circles which are exterior to A cannot form part of the boundary of R_0 , it follows that R_0 is bounded by a finite number of sides.

The group is then generated by a finite number of transformations.

Classification of Function Groups.—We shall separate the function groups into three major classes.

1. *Elementary Groups.*—These consist of the finite groups and the groups with one or two limit points.

2. *Fuchsian Groups.*—A group is called "Fuchsian" if its transformations have a common fixed circle and if each transformation carries the interior of the fixed circle into itself.

3. *Kleinian Groups.*—A function group is called "Kleinian" if it does not belong to one of the preceding classes.

Examples of elementary groups are the group discussed in Sec. 22 and the cyclic groups of Sec. 24.

There is a certain amount of overlapping between (1) and (2). Certain of the elementary groups possess fixed circles; for example, the non-loxodromic cyclic groups.

Most of the combination function groups are Kleinian groups.

CHAPTER III

FUCHSIAN GROUPS

29. *The Transformations.*—As defined in the preceding section, a Fuchsian group is a properly discontinuous group each of whose transformations carries a certain circle into itself and carries each of the parts into which the circle divides the plane into itself. The common fixed circle will be called the "principal circle." A point within the principal circle is carried into an interior point; an exterior point is carried into an exterior point. If the principal circle is a straight line, a point on one side of the line is carried into a point on the same side of the line.

Consider now the kinds of transformations that can belong to a Fuchsian group. We found in Sec. 9 that the only loxodromic transformations which have fixed circles carry the interior of each fixed circle into its exterior. Hence, there can be no loxodromic transformation. Referring to Secs. 7, 8, 10, where we studied the fixed circles of the non-loxodromic transformations, and to the accompanying figures, we see that a transformation of the group, other than the identical transformation, must be one of the following kinds:

(a) A hyperbolic transformation with its fixed points on the principal circle.

(b) An elliptic transformation with its fixed points inverse to one another with respect to the principal circle.

(c) A parabolic transformation with its fixed point on the principal circle and with its fixed straight line tangent to the principal circle.

Conversely, a transformation of any one of these three kinds carries the circle into itself.

An application of Theorem 20, Sec. 11(c), gives the following result:

THEOREM 1.—*The isometric circles of the transformations of a Fuchsian group are orthogonal to the principal circle.*

30. *The Limit Points.*—Since the principal circle is transformed into itself, it constitutes a set to which Theorem 5 of Sec. 18 applies. Hence, we have the following theorem:

THEOREM 2.—*The limit points of a Fuchsian group lie on the principal circle.*

This theorem can also be proved directly from Theorem 1. For, the requirements that a limit point shall have isometric circles of arbitrarily small radius in its neighborhood and that these circles shall be orthogonal to the principal circle can be met only if the limit point is on the principal circle.

THEOREM 3.—*If there are more than two limit points, either (1) the set of limit points consists of all points of the principal circle; or (2) the set of limit points is a perfect set which is nowhere dense on the principal circle.*

We have already found (Sec. 18, Theorem 4) that if there are more than two limit points the limit points form a perfect set of points. We must show further that unless every point of the principal circle is a limit point, then the limit points form a set which does not contain all the points of any arc of the principal circle.

Let z_0 be a point lying on the principal circle and which is not a limit point. Then the points in the neighborhood of z_0 are also ordinary points. In particular, the points on a suitably small arc h of the principal circle passing through z_0 are ordinary points.

Let P be a limit point. Then, applying Theorem 2, Sec. 18, there are points congruent to points of h in the neighborhood of P . These points are ordinary points and lie on the principal circle. Since there are ordinary points on the principal circle in the neighborhood of a limit point, it follows that the set of limit points is nowhere dense on the principal circle.

It is known from the theory of perfect sets that the set of limit points in (2) can be formed by the removal from the principal circle of an infinite number of open arcs, $h_1, h_2, \dots$. These arcs do not overlap and have no common end points. Further between any two arcs lie infinitely many others. The perfect set consists of the points that remain after the arcs have been removed.

On the basis of the preceding theorem we shall classify Fuchsian groups as follows:

(a) *Fuchsian groups of the first kind*, or groups for which every point of the principal circle is a limit point.

(b) *Fuchsian groups of the second kind*, or groups whose limit points are nowhere dense on the principal circle.

To (b) belong the groups whose limit points form a non-dense perfect set and also the elementary Fuchsian groups, where the number of limit points is finite.

In (a) the limit points separate the plane into two regions each of which is carried into itself. In (b) the ordinary points form a single connected region.

We shall find that groups of both kinds exist, and that the two kinds of groups have strikingly different properties.

31. The Region R and the Region R_0 .—When we make an inversion in a circle, any orthogonal circle is carried into itself and the interior of the orthogonal circle is carried into itself. Hence, from Theorem 1, if we make an inversion in the principal circle, each isometric circle is carried into itself and its interior and exterior go, respectively, into its interior and exterior. A point of R , being exterior to all the isometric circles, is carried into another point of R . Hence, we have the following result:

THEOREM 4.—*An inversion in the principal circle carries the region R into itself.*

In this inversion, infinity is carried into the center of the principal circle, providing, as we shall assume, that the principal circle is not a straight line. Since the points in the neighborhood of infinity lie in R , it follows that points in the neighborhood of the center of the principal circle lie in R .

The principal circle divides R into two parts which may or may not be connected with one another along the principal circle.

We shall designate by R_0 the part of R lying within the principal circle, by R_0' the part of R lying without the principal circle.

R_0' is the inverse of R_0 in the principal circle; its sides and vertices are the inverses of the sides and vertices of R_0 . The sides which bound R_0' lie on the same isometric circles as the sides which bound R_0 . Corresponding sides are connected by the same transformations.

Let a transformation of the group be made. R_0 is carried into a region in the interior of the principal circle and R_0' into a region on the exterior. As a consequence of Theorem 10, Sec. 5, we can state that the two transformed regions are inverse with respect to the principal circle.

It follows from the preceding remarks that, in a study of the fundamental region, of its sides and vertices, of its congruent regions, and the like, it will suffice to study the region R_0 within the principal circle. An inversion in the principal circle will then furnish the corresponding results for R_0' . We shall, therefore, limit our study to R_0 .

THEOREM 5.—*The region R_0 is simply connected.*

It is clear that the region R_0 is connected. A straight line segment from the center of the principal circle to any point of R_0 lies entirely within R_0 . Thus any two points of R_0 can be joined by a curve lying in R_0 ; for example, by combining the line segments joining each to the center.

R_0 is simply connected if any closed curve in R_0 can be shrunk continuously to an interior point without crossing the boundary. This can be done by the simple process of moving the points of the curve continuously along radii to the center.

The main facts concerning the sides of R_0 are summarized in the following theorem:

THEOREM 6.—*The sides of R_0 are circular arcs orthogonal to the principal circle. These arcs are congruent in pairs. Two congruent sides are equal in length and congruent points thereon are equidistant from the center of the principal circle.*

That the sides are orthogonal to the principal circle follows from Theorem 1. That the sides are arranged in congruent pairs which are equal in length follows from the fact (Sec. 21, Theorem 10) that the sides of R are so arranged. A side of R_0 is a side, or a portion of a side, of R , and is equal in length to its congruent side. This congruent arc lies also within the principal circle and is a side of R_0 .

That congruent points on two congruent sides of R_0 are equidistant from the center of the principal circle is a consequence of Theorem 21, Sec. 11.

Attention should be called to the fact that an arc of the principal circle along which R_0 and R_0' are adjacent is not considered as a side of R_0 . The term "side" is here limited to arcs of isometric circles.

THEOREM 7.—*Any closed region lying entirely within the principal circle is covered by a finite number of transforms of R_0 . These regions fit together without lacunæ.*

This theorem follows from the fact that a region of the kind specified is covered without lacunæ by a finite number of transforms of R (Sec. 20, Theorem 8) and from the further fact that R_0 contains all the points of R whose transforms lie within the principal circle.

THEOREM 8.—*The transforms of R_0 fill up, without lacunæ, the whole interior of the principal circle. They cluster in infinite number about each limit point of the group.*

The first part of this theorem is a consequence of Theorem 7. A circle Q concentric with the principal circle and of smaller radius is covered by R_0 and a finite number of regions congruent to R_0 . By taking the radius of Q near enough to that of the principal circle we can enclose in Q any given point interior to the principal circle. It follows that the interior of the principal circle is completely covered. The second part of the theorem follows from Theorem 9, Sec. 20.

THEOREM 9.—*An interior point of R_0 is nearer the center of the principal circle than any point congruent to it.*

An interior point of R_0 is outside the isometric circle of any transformation by which it is carried into a congruent point. The theorem then follows from Theorem 21, Sec. 11.

32. Generating Transformations.

THEOREM 10.—*The transformations by which the sides of R_0 are congruent form a set of generating transformations for the group.*

If the Fuchsian group is of the first kind, the theorem follows directly from Theorem 19, Sec. 28; for R_0 is that part of R lying in the interior of the principal circle, which may be taken as the region Σ of Sec. 28. If the group is of the second kind, Σ consists of all ordinary points in the plane, and the fundamental region of that theorem is R itself. But the sides of R which lie outside the principal circle are congruent by the same transformations that connect the sides of R_0 and so supply no new generating transformations.

Let Q be a circle lying within and concentric with the principal circle, and let us consider the network of regions by which Q is covered. We shall prove the following theorem:

THEOREM 11.—*A circle Q concentric with the principal circle and of smaller radius is completely covered by R_0 and by regions which are congruent to R_0 by transformations formed by combining those generating transformations which connect sides of R_0 lying wholly or in part in Q .*

The network of regions covering Q is found by adjoining regions congruent to R_0 along the sides of R_0 lying in Q , adjoining regions along such sides of these transformed regions as lie in Q , and so on (Sec. 23). In this process, we employ only those generating transformations connecting sides of R_0 which lie in Q or are congruent to sides lying in Q . For example, if one of these regions, $T(R_0)$, has a side l in Q which is congruent to l_k of R_0 , then $TT_k(R_0)$ is the region adjacent to $T(R_0)$ along l . Hence, at each step we introduce only the generating transformations stated.

Now, when R_0 is carried into a congruent region, the distance of an interior point from the center of the principal circle is increased (Theorem 9); hence, the distance of a boundary point is not decreased. Then all transforms of all sides of R_0 which lie outside Q are themselves outside Q ; and the

transformations connecting these sides play no part in the formation of the network. This establishes the theorem.

THEOREM 12.—Let S be a transformation of the group; let P be the point on the isometric circle of S which is nearest the center of the principal circle; and let Q be the circle through P concentric with the principal circle. Then S can be expressed as a combination of those generating transformations which connect sides of R_0 which have at least one interior or end point within or on Q .

The theorem is so worded as to include not only sides lying wholly or in part in Q but also sides which touch Q or have an end point lying on Q . Since there is but a finite number of sides in the neighborhood of Q we can draw a slightly larger circle Q' , concentric with Q , which contains no new sides of R_0 . Q' contains P on I_s and the congruent point P' on $I_{s'}$, since P' and P are equidistant from the center of the principal circle.

It follows from the preceding theorem that Q' is completely covered by R_s and the regions congruent to R_s by transformations of the kind mentioned in the present theorem. These regions cover the neighborhoods of P and P' . S carries interior points of a region $T(R_s)$ in the neighborhood of P into interior points of some region $T'(R_s)$ in the neighborhood of P' , where both T and T' are combinations of generating transformations of the kind specified in the theorem. Then $ST(R_s)$ and $T'(R_s)$ overlap, whence (Theorem 1, Sec. 16) $ST = T'$; and $S = TT'^{-1}$. This is a combination of the kind required; and the theorem is established.

33. The Cycles.—Let $A_1, A_2, \dots, A_m$ be the vertices of an ordinary cycle of R , arranged in order as in Sec. 26. Then A_k and A_{k+1} are congruent points at the ends of congruent sides l_k, l_{k+1} of R . We need consider only vertices of R_0 , those of R_0' being got by an inversion. It follows from Theorem 6 that A_k and A_{k+1} are equidistant from the center of the principal circle.

The congruent points of a parabolic cycle are limit points and lie on the principal circle. We have, then, the following theorem:

THEOREM 13.—The congruent vertices of an ordinary cycle lie on a circle concentric with the principal circle. The points of a parabolic cycle lie on the principal circle.

There arises the question whether there can be an ordinary cycle with vertices lying on the principal circle. Such cycles do exist for certain groups. Two isometric circles which meet on the principal circle, being both orthogonal to the principal circle, are tangent. Hence, the angle at each vertex is either 0 or π .

In Fig. 24 is shown the region R for a combination group arising from two hyperbolic cyclic groups. Two of the circles are made tangent. The order of the sides and vertices, according to the scheme of Sec. 26 is indicated. The congruent vertices are in order $A_1, A_2, A_3 (= A_1), A_4$.

This situation can arise only if the sum of the angles of the cycle is 2π . Otherwise, the transformation $S = T_m \cdots T_1$, which carries A_1 into itself, is an elliptic transformation and (Sec. 29 (b)) the fixed point A_1 does not lie on the principal circle. In the figure, the angles at A_2 and A_4 are each π and those at A_1 and A_3 are zero.

The cycle in this problem is not an essential one and we can remove it entirely if we wish. We can replace any part of R by a congruent part and still have a fundamental region. Let the sides l_1 and l_2 be slightly deformed in the neighborhood of A_2 so that A_2 becomes an interior point. The part of R which is removed in the neighborhood of A_2 can be replaced by the congruent part lying within I_1' in the neighborhood of A_1 . Then, l_1 and l_2 are displaced slightly to the right and the sides no longer touch.

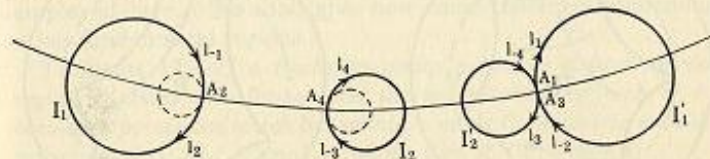


FIG. 24.

It is not difficult to see that this can be done in the most general case. In such a cycle, two of the vertices, as A_2 and A_4 , have the angle π , and a finite number have the angle zero. By deforming the sides at one of the former vertices we can remove one of the points of tangency. By deforming the sides which now meet at the latter point we remove another point of tangency. By successive steps we can remove the remainder.

34. Fuchsian Groups of the First and Second Kinds.—Whether a Fuchsian group is of the first or the second kind (Sec. 30) depends upon the region R , as stated in the following theorem:

THEOREM 14.—If R contains on its interior a point of the principal circle, the group is of the second kind; if not, it is of the first kind.

The first part of the theorem is evident. If a point of the principal circle is on the interior of R , it is an ordinary point. Not all points of the principal circle are limit points; and the group is of the second kind.

To complete the proof we show that if the group is of the second kind, R contains a point of the principal circle on its interior. Let z_0 be an ordinary point on the principal circle. Then a circle Q can be drawn about z_0 such that all points within and on Q can be covered by a finite number of transforms

of R (Theorem 8, Sec. 20), $R_1, \dots, R_n$, say. Each of these regions has a finite number of sides lying in Q . For, an infinite number of sides would have a cluster point in or on Q ; and this cluster point, being the transform of a point of R at which infinitely many isometric circles cluster, would itself be a limit point, which is impossible. Further, the sides of these regions, being congruent to the sides of R , are orthogonal to the principal circle. Let h be the arc of the principal circle lying in Q . Then, with the exception of the finite number of points where the sides of $R_1, R_2, \dots, R_n$ meet the principal circle, each point of h is interior

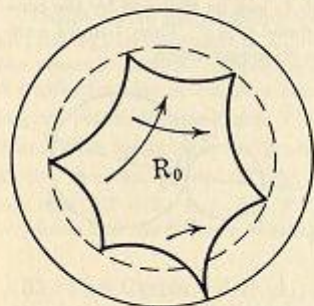


FIG. 25.

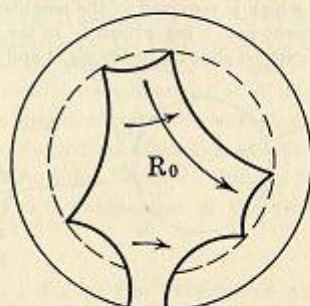


FIG. 26.

to one of the regions. Let z_1 be a point of h interior to R_k . Making the transformation which carries R_k into R , z_1 is carried into a point which is interior to R and which lies on the principal circle. The existence of this point establishes the theorem.

If R contains a point of the principal circle on its interior, it contains all points of an arc of the principal circle passing through the point. In the group of the second kind, then, R contains one or more arcs of the principal circle on its interior. The two regions R_0 and R_0' abut along these arcs.

The region R_0 for the two types of groups has the character illustrated in the accompanying figures. In the group of the first kind (Fig. 25) R_0 either lies, together with its boundary, within the principal circle; or, if there are points of the boundary of R_0 on the principal circle they are limit points of the group with sides of R_0 in the neighborhood. In the group of the second kind (Fig. 26) R_0 abuts on the principal circle along one or more arcs, and R contains these arcs on its interior.

THEOREM 15.— R_0 constitutes a fundamental region for the Fuchsian group of the first kind.

This theorem is a consequence of Theorem 18, Sec. 28.

THEOREM 16.—If the boundary of R_0 lies within the principal circle, or if the only boundary points on the principal circle are parabolic points, then R_0 has a finite number of sides. The group is then generated by a finite number of transformations.

If the only boundary points of R_0 which lie on the principal circle are parabolic points, the same is true of the boundary points of R_0' . The first part of the theorem then follows from Theorem 17, Sec. 27. The latter is then a consequence of Theorem 10.

In our later work, particularly in the study of uniformization, we shall come upon Fuchsian groups of the first kind where the fundamental region is found in quite other ways than that employed here. We shall give now some theorems concerning other fundamental regions.

THEOREM 17.—If a Fuchsian group possesses a fundamental region F whose transforms cover the neighborhood of each of its boundary points and which lies within a circle Q concentric with the principal circle and of smaller radius, then R_0 lies within Q .

A finite number of transforms of R_0 cover F completely. When we carry the portion of F in each region R_i into R_0 by means of the transformation which carries R_i into R_0 these transformed parts of F fill R_0 completely without lacunæ, as shown in the proof of Theorem 20, Sec. 28. Then, R_0 is in Q ; for on transforming the parts of F into congruent parts in R_0 the distance of no point from the center of the principal circle is increased (Theorem 9).

THEOREM 18.—The transforms of F fill up, without overlapping and without lacunæ, the whole interior of the principal circle.

The transforms of F cover R_0 completely. The transforms of F , then, cover all transforms of R_0 . These fill up the interior of the principal circle. There can be no overlapping of the transforms of F since no two points of F are congruent.

THEOREM 19.—Of the fundamental regions lying within the principal circle, R_0 has the maximum area.

A fundamental region different from R_0 is formed by replacing parts of R_0 by congruent parts. Since R_0 is exterior to all isometric circles, a shift of any part of R_0 to a congruent position effects a diminution of area.

35. Fixed Points at Infinity. Extension of the Method.—If a transformation has a fixed point at infinity, either there is no

isometric circle or all circles are isometric. In the study of a group, we have supposed hitherto that it has been so transformed that there are no transformations with ∞ as a fixed point. This involves, of course, no loss of generality. There is, however, sometimes a distinct loss of simplicity in the definition of the group. In the present section, we shall discuss briefly the formation of the fundamental region for a properly discontinuous group certain of whose transformations have fixed points at infinity.

THEOREM 20.—*All the transformations of a given group which leave some configuration unaltered constitute a subgroup of the given group.*

For, the succession of two transformations which leave the configuration unaltered and also the inverse of any, leave the configuration unaltered, and they belong to the given group.

Taking the point at infinity as the configuration of the theorem, we have the following corollary:

COROLLARY.—*All those transformations of a given group which leave the point at infinity fixed constitute a subgroup of the given group.*

We shall represent by $U_0 (=1)$, U_1 , U_2 , . . . the transformations with ∞ as fixed point and shall call the subgroup which they form Γ_∞ .

All transformations other than those of Γ_∞ possess isometric circles. It is no longer necessarily true that the radii of these circles are bounded, that their centers lie in a finite region, or that the centers are distinct. Concerning these isometric circles we shall prove the following theorem:

THEOREM 21.—*A transformation of the group Γ_∞ carries an isometric circle into an isometric circle.*

Let I_1 be the isometric circle of T , and let a transformation U be applied. We shall show that $U(I_1)$ is the isometric circle of the transformation $S = UTU^{-1}$. U has the form

$$z' = U(z) = Kz + b,$$

where K is the multiplier, as we found in Secs. 6 and 10, Equations (31) and (39'). Since $U'(z) = K$, the transformation multiplies all lengths by $|K|$; the transformation U^{-1} multiplies all lengths by $1/|K|$.

Let P be a point on the circle $U(I_1)$ and let P be transformed by S . U^{-1} carries P to P' , a point on I_1 , lengths in the neighborhood being multiplied by $1/|K|$. T carries P' into P'' without alteration of lengths. U multiplies lengths in the neighborhood of P''

by $|K|$. The result is that UTU^{-1} has not altered lengths in the neighborhood of P ; hence P is on the isometric circle of I_1 .

Let us now construct a fundamental region F for the subgroup Γ_∞ . The existence of F follows (after a preliminary transformation of the group Γ_∞) from the general theory of the preceding chapter—sides congruent in pairs, ordinary vertices arranged in cycles, etc. In many cases, however, as in Figs. 11 to 14, we know a fundamental region already, and we do not need to employ the general theory.

The principal theorem of the present section will now be proved.

THEOREM 22.—*Let F be a fundamental region for the subgroup Γ_∞ , where the sides of F are congruent in pairs and the transforms of F cover the finite plane,¹ and let R consist of all that part of F which is exterior to all isometric circles of the group. Then R is a fundamental region for the group.*

A transformation of Γ_∞ carries a point of R into a point of a region congruent to F and hence outside R . Any other transformation of the group carries a point of R into the interior of an isometric circle and hence outside R . So no two points of R are congruent.

On the assumption that the sides of F are congruent in pairs, the sides of R are congruent in pairs. That part of a side of F which is exterior to all isometric circles, and so is also a side of R , is, from Theorem 21, congruent to a side which is also exterior to all isometric circles. An ordinary boundary point P of R lying on an isometric circle I_1 is carried by T into a point P' on I_1' . We can show in the usual way that P' is interior to no isometric circle. If P' lies in F , it is a boundary point of R ; if not, it lies in a congruent region $U(F)$, and the congruent point $P'' = U^{-1}(P')$ lies in F and is a boundary point of R . It follows that the sides of R which lie on isometric circles are congruent in pairs.

A region about a boundary point of R on a side l_1 contains points congruent to points of R in the neighborhood of the congruent side l_{-1} ; the neighborhood of a limit point contains points congruent to points of R by Theorem 2, Sec. 18. Hence, R is a fundamental region.

The further properties of R , such as the arrangement of the vertices in cycles, the theorems on the generating transformations,

¹ The fundamental regions of Figs. 11 to 14 have these properties. That the most general subgroup Γ_∞ has a fundamental region whose transforms cover the whole finite plane will appear in Chap. VI.

and the like, are established as in our previous treatment, and will not be repeated here.

36. Examples. The Group of the Anharmonic Ratios.—This group (Sec. 13 (2)) contains, in addition to the identical transformation, the following transformations:

$$U = 1 - z, \quad T_1 = \frac{1}{z}, \quad T_2 = \frac{1}{1-z}, \\ T_3 = \frac{z-1}{z}, \quad T_4 = \frac{z}{z-1}.$$

Here U is a rotation through the angle π about the point $\frac{1}{2}$. A fundamental region for the subgroup Γ_u is the half plane bounded by any straight line through $\frac{1}{2}$; for instance, the upper half plane.

The isometric circles of the remaining transformations are

$$I_1, I_3: |z| = 1; \quad I_2, I_4: |z-1| = 1.$$

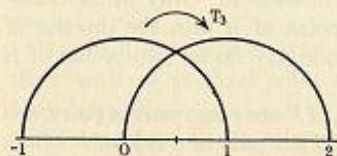


FIG. 27.

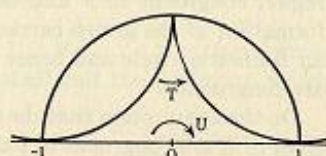


FIG. 28.

We have, then, as a fundamental region R for the group that part of the upper half plane lying outside these circles (Fig. 27).

The rectilinear sides of R are congruent by U , the circular sides by T_2 . U and T_2 are then generating transformations. From the two cycles of angles $2\pi/3$ and π , we have the relations $T_2^3 = 1$ and $(UT_2)^2 = 1$.

A Group with Q_0 as Principal Circle.—As a further example, we shall consider the group mentioned in Sec. 13 (7):

$$z' = \frac{az + \bar{c}}{cz + \bar{a}}, \quad a\bar{a} - c\bar{c} = 1,$$

where a and c are complex integers. This is a Fuchsian group with Q_0 as principal circle.

When $c = 0$, we have, $a\bar{a} = 1$; whence, $a = \pm 1$ or $\pm i$. This gives for Γ_u the two transformations,

$$z' = z, \quad z' = U(z) = \frac{iz}{-i} = -z.$$

U is a rotation about the origin through the angle π . A fundamental region for Γ_u is the half plane above the real axis.

Among the remaining transformations, consider, first, those with the largest isometric circles. The smallest absolute value of c occurs when $c = \pm 1$ or $\pm i$. Then $a\bar{a} = 1 + c\bar{c} = 2$; and $a = \pm 1 \pm i$. This gives for the center $-\bar{a}/c$, the possible positions $1+i$, $1-i$, $-1+i$, $-1-i$. Only two circles with these centers and of unit radius lie in the upper half plane. They are the isometric circles of

$$T = \frac{(1+i)z+1}{z+1-i}$$

and its inverse. They are shown in Fig. 28.

We shall now show that each other isometric circle in the upper half plane is contained in one of these. Each such circle is orthogonal to Q_0 and of radius less than 1. Unless it lies within or is tangent internally to one of the two circles already drawn, it must contain one of the points ± 1 or i on its interior. Its isometric circle is

$$|cz + \bar{a}| = 1, \quad |c| > 1.$$

If one of the three points lies within this circle, we have

$$|\pm c + \bar{a}| < 1, \quad \text{or} \quad |ic + \bar{a}| < 1.$$

Since the term whose absolute value appears in the first member is an integer, its absolute value can be less than 1 only if it is zero. Then $|c| = |a|$, and $a\bar{a} - c\bar{c} = 0$, which is impossible.

R , then, is that part of the upper half plane which is exterior to the two circles of Fig. 28. R contains no points of Q_0 ; so the group is of the first kind. All points of Q_0 are limit points. R consists of two parts: R_0 within Q_0 , and R'_0 , the inverse of R_0 in Q_0 . The transforms of R_0 fill up the whole interior of Q_0 . U and T are generating transformations.

There are three cycles. The origin constitutes an ordinary cycle of angle π , whence $U^2 = 1$. The point i constitutes a parabolic cycle. The points 1 and -1 form a second parabolic cycle.

37. The Modular Group.—This group (Sec. 13 (5)) consists of the transformations

$$z' = \frac{az + b}{cz + d}, \quad ad - bc = 1,$$

where a, b, c, d , are real integers. The real axis is a fixed circle. Whether the upper half plane is carried into itself or into the

lower half plane can be determined from a consideration of one point. When $z = i$, we find $z' = [ac + bd + (ad - bc)i]/(c^2 + d^2)$. The imaginary part of this quantity is positive; so z' is also in the upper half plane. The group is thus a Fuchsian group.

When $c = 0$, we have $ad = 1$. Then $a = d = \pm 1$, and b can have any integral value. This gives for the subgroup Γ_n (Sec. 35) the set of transformations

$$z' = z + n,$$

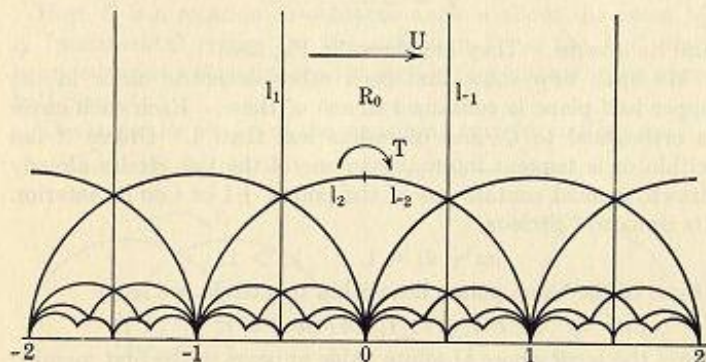


FIG. 29.

where n is any integer. This is the group of translations whose fundamental region—setting $\omega = 1$ —is shown in Fig. 12. We shall take as fundamental region the strip enclosed by lines perpendicular to the real axis through the points $\pm \frac{1}{2}$. The subgroup is generated by the transformation

$$z' = U(z) = z + 1.$$

Consider the largest isometric circles. If $c = \pm 1$, we have $ad \mp b = 1$. For any integral value of a and d , we can determine from this equation an integral value of b . It follows that the center, $\mp d$, of the isometric circle $|\pm z + d| = 1$ can be any integer. These circles, then, are the unit circles with centers at the real integers. They are the large circles in Fig. 29.

These isometric circles enclose all points within a distance of $\frac{1}{2}\sqrt{3}$ of the real axis. For any other transformation $|c| \geq 2$; and the isometric circle is of radius not exceeding $\frac{1}{2}$. As the center is on the real axis, such a circle lies in the space enclosed

by the unit isometric circles. These smaller circles, then, can form no part of the boundary of R .

The region R , lying in the period strip previously selected and exterior to the isometric circles, is bounded by the unit circle with center at the origin. It consists of the region R_0 shown in the figure together with the reflection of R_0 in the real axis. The circular boundary of R_0 consists of two sides l_2 and l_{-2} which are congruent by

$$z' = T(z) = -\frac{1}{z},$$

an elliptic transformation with fixed points $\pm i$. U and T are generating transformations for the group.

There are three cycles. The point at infinity, to which R_0 extends, is a parabolic point and constitutes a cycle. The point i constitutes a cycle of angle π ; whence $T^2 = 1$. The remaining two vertices, namely, $\pm \frac{1}{2} + \frac{1}{2}\sqrt{3}i$, constitute a cycle of angle $2\pi/3$.

To get the relation connecting U and T which arises from the last cycle, we proceed as in Sec. 26. Starting from the right-hand vertex and the side l_{-1} , beginning there, we get the transformation U^{-1} then T before returning. Then,

$$S = TU^{-1} = -\frac{1}{z-1}$$

is a transform of period three which carries the right-hand vertex into itself. The desired relation is $S^3 = 1$ or $TU^{-1}TU^{-1}TU^{-1} = 1$. This can also be written, if we take the inverse and use the fact that $T^{-1} = T$, in the form

$$UTUTUT = 1.$$

The transforms of R_0 cover the whole upper half plane. A number of the regions congruent to R_0 are shown drawn to scale in the figure. The regions cluster in infinite number about each point of the real axis.

38. Some Subgroups of the Modular Group.—As a further exemplification of the method of forming the fundamental region, we shall now consider a particular set of the great variety of subgroups contained within the modular group. Let n be an integer greater than 1, and consider all those transformations of

the modular group for which b and c are divisible by n . We then have all transformations of the form

$$T = \frac{az + nb'}{nc'z + d}, \quad ad - n^2b'c' = 1,$$

where a, b', c', d are real integers.

We prove first that these transformations form a group. The inverse, $T^{-1} = (-dz + nb')/(nc'z - a)$ is of the same form. Let $S = (\alpha z + n\beta')/(n\gamma'z + \delta)$ be a second transformation of the set, then

$$ST = \frac{(\alpha a + n^2\beta'c')z + n(\alpha b' + \beta'd)}{n(\gamma'a + \delta c')z + n^2\gamma'b' + \delta d}.$$

Since $n(\alpha b' + \beta'd)$ and $n(\gamma'a + \delta c')$ are divisible by n , this belongs to the set. Thus both group properties are satisfied.

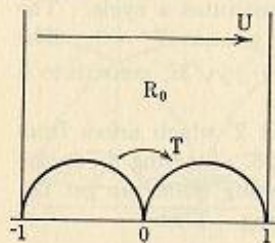


FIG. 30.

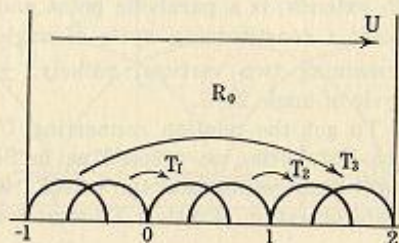


FIG. 31.

Setting $c' = 0$, we get for the group Γ_n the set of transformations

$$U_m = z + mn,$$

where m is any integer. Γ_n is generated by $U = z + n$.

In the accompanying figures, fundamental regions for two cases are shown. The reasoning follows the lines of the preceding sections and is left to the reader.

Figure 30 shows R_0 for the case $n = 2$. The group is generated by the two transformations

$$U = z + 2, \quad T = \frac{z}{2z + 1}.$$

There are three parabolic cycles.

Figure 31 is for the case $n = 3$. The generating transformations are

$$U = z + 3, \quad T_1 = \frac{z}{3z + 1}, \quad T_2 = \frac{4z - 3}{3z - 2}, \quad T_3 = \frac{5z + 3}{3z + 2}.$$

There are four parabolic cycles and one ordinary cycle of angle 2π .

CHAPTER IV

AUTOMORPHIC FUNCTIONS

39. The Concept of the Automorphic Function.—Automorphic functions are the generalization of the circular, hyperbolic, elliptic, and certain other functions of elementary analysis. A circular function, as $\sin z$, has the property that it is unchanged in value if z is replaced by $z + 2m\pi$, where m is any integer; that is, the function is unaltered in value if z be subjected to a transformation of the group $z' = z + 2m\pi$. A hyperbolic function, such as $\sinh z$, is unchanged in value if z be subjected to a transformation of the group $z' = z + 2m\pi i$. An elliptic function, as the Weierstrassian function $\wp(z)$, retains its value under transformations of a group of the form $z' = z + m\omega + m'\omega'$.

The automorphic function is an extension of this concept to the more general properly discontinuous group. Roughly speaking, a function is automorphic with respect to such a group if it has the same value at congruent points. We shall lay down a more precise definition.

By the *domain of existence* of a single-valued analytic function $f(z)$ we shall mean the set of points at which $f(z)$ is analytic or has poles. The domain of existence is a connected region consisting entirely of interior points,—a two dimensional continuum.

DEFINITION.—A function $f(z)$ will be said to be automorphic with respect to a group of linear transformations $T_1, T_2, \dots$ provided

1. $f(z)$ is a single-valued analytic function.
2. If z lies in the domain of existence of $f(z)$ so also shall $T_n(z)$.
3. $f[T_n(z)] = f(z)$.

Because of the first condition, the functions here defined might more properly be called "single-valued" automorphic functions. There exist many-valued analytic functions satisfying conditions (2) and (3). But we shall be concerned altogether with single-valued functions; and to avoid repeatedly calling attention to this fact we shall make single-valuedness a part of the definition.

We note that there are no functions, other than constants, which are automorphic, according to this definition, with respect to a continuous or an improperly discontinuous group. For, let $F(z)$ be such a function and let z_0 be a point at which the function is analytic. There are infinitely many points in the neighborhood of z_0 which are congruent to z_0 . At each of these points $F(z) = F(z_0)$. It is a well-known fact that a function can take on the same value at an infinite number of points in the neighborhood of a point at which it is analytic only if it is constant.

It is observed from the definition, that if $f(z)$ is automorphic with respect to a group it is automorphic with respect to any subgroup.

In showing that a function is automorphic with respect to a group, it is not necessary to investigate conditions (2) and (3) for all points of the domain of existence of the function nor for all transformations of the group. As the following theorems show, it suffices to establish the conditions for some small region and for the generating transformations of the group.

THEOREM 1.—Let $f(z)$ be a single-valued function analytic at z_0 . Let $T(z_0)$, where T is a linear transformation, lie in the domain of existence of the function; and let

$$f[T(z)] = f(z) \quad (1)$$

be valid in the neighborhood of z_1 . Then, if z is any point in the domain of existence of the function, so also is $T(z)$, and (1) holds throughout the whole domain of existence.

The transformation T carries the domain of existence of $f(z)$ into itself.

This theorem is an immediate consequence of the principle of analytic continuation. The two functions of z that appear in (1) are identical in a region surrounding z_1 ; they are, therefore, identical in any region to which either can be continued analytically. Let z_1 be a point at which $f(z)$ is analytic. Then $f(z)$ can be continued analytically from z_1 throughout a suitable region S surrounding z_1 . Then in S $f(T(z))$ is analytic and (1) holds. That is, $f(z)$ is analytic in the neighborhood of $z_1' = T(z_1)$ and $f(z_1') = f(z_1)$.

In the neighborhood of a pole z_2 , $f(z)$ is analytic and $f(z)$ becomes infinite as z approaches z_2 . Then $f(z)$ is analytic in the neighborhood of $z_2' = T(z_2)$ and becomes infinite as z_2' is approached.

T carries any point of the domain of existence of $f(z)$ into another such point. The same is true of T^{-1} ; since $f[T^{-1}(z)] = f(z)$ holds in the neighborhood of $z_1' = T(z_0)$. Then T carries no point without the domain of existence into the domain. Hence, T carries the domain of existence into itself.

THEOREM 2.—If $f(z)$ is a single-valued analytic function and if

$$f[T_1(z)] = f(z), \quad f[T_2(z)] = f(z), \quad \dots,$$

then $f(z)$ is automorphic with respect to the group generated by $T_1, T_2, \dots$.

Each transformation of the group carries the domain of existence of $f(z)$ into itself.

The group is formed by constructing all possible products by means of $T_1, T_2, \dots$ and their inverses. It is clear that $f(z)$ is unaltered when z is replaced by $T_1^{-1}(z), T_2^{-1}(z), \dots$. Since any product can be formed by repeatedly combining transformations two at a time it suffices, to prove the first part of the theorem, to point out that if $f[S(z)] = f(z), f[T(z)] = f(z)$, then $f[ST(z)] = f(z)$. But if z is in the domain of existence of the function so is $T(z)$, and hence $ST(z)$, and we have $f[ST(z)] = f(T(z)) = f(z)$.

The latter part of the theorem is an application of the latter part of Theorem 1.

Consider, as an example, the function $\cos z$. We have $\cos(z + 2\pi) = \cos z$ and $\cos(-z) = \cos z$. Then $\cos z$ is automorphic with respect to the group generated by $z' = z + 2\pi$ and $z' = -z$.

The existence of a non-constant single-valued analytic function which is unaltered when a set of linear transformations is applied to the independent variable is sufficient to show that the group generated by the transformations is properly discontinuous. Thus, the group mentioned in the preceding paragraph is necessarily properly discontinuous.

THEOREM 3.—The domain of existence of an automorphic function extends into the neighborhood of every limit point of the group.

For, in the neighborhood of a limit point lie points congruent to points in the domain of existence of the function. These points belong to the domain of existence of the function.

THEOREM 4.—If the automorphic function is not a constant, each limit point of the group is an essential singularity of the function.

In the neighborhood of a point at which a function is analytic or has a pole, the function can take on any value only a finite number of times. In the neighborhood of a limit point there is an infinite number of congruent points at which the function takes on the same value. Hence, the limit point is an essential singularity.

COROLLARY.—All points of the domain of existence of a non-constant automorphic function are ordinary points.

It is not true, of course, that all points on the boundary of the domain of existence of the function are necessarily limit points. The limit points lie on the boundary of the domain, but there may be further boundary points. For example, the function $e^{\frac{1}{z}}$ is automorphic with respect to the group $z' = z + m\omega + m'\omega'$. Its domain of existence consists of all points except the point ∞ (the only limit point) and the points $m\omega + m'\omega'$. At these latter points $\mathfrak{P}(z)$ has poles and the function has essential singularities.

It results from the preceding discussion that not all properly discontinuous groups have non-constant automorphic functions.

THEOREM 5.—If a group possesses a non-constant automorphic function, it is a function group.

Let a group have a non-constant automorphic function existing in a domain S . Let Σ be the part of the plane, bounded by limit points, in which S lies. Σ consists of all ordinary points which can be joined to a point of S by curves not passing through limit points. Any point z of Σ and a curve C joining it to a point of S are carried by any transformation T of the group into a point z' and a curve C' joining z' to a point of S , where C' consists of ordinary points. Then z' belongs to Σ ; whence Σ is carried into itself. The group is, therefore, a function group.

We shall find in the following chapter that every function group possesses non-constant automorphic functions.

40. Simple Automorphic Functions.—In the present chapter, and in our later work, we shall have much to do with automorphic functions of a somewhat restricted kind. We shall impose restrictions both on the character of the function and on the group with respect to which it is automorphic.

Let $f(z)$ have no essential singularity at an ordinary point of the group. Then the domain of existence of the function, provided it is not a constant, is one of the regions Σ into which the limit points of the group separate the plane.

That part R_0 of R which lies in Σ is a fundamental region for the group. We shall require that R_0 have a finite number of sides.

If R_0 possesses one or more parabolic points we shall impose a further condition on the function. As z approaches a parabolic point P from the region R_0 let the function approach a definite value, finite or infinite; that is,

$$\lim_{z \rightarrow P} f(z) = C, \text{ or } \infty,$$

where z is restricted, in its approach, to lie within or on the boundary of R_0 .¹

To avoid long circumlocutions in the statement of theorems we shall call such a function a "simple automorphic function." A simple automorphic function then (1) belongs to a function group such that R_0 has a finite number of sides; (2) has the domain of existence Σ , provided it is non-constant; and (3) behaves in the manner specified at the parabolic points, if any.

If the group is finite, Σ consists of the whole plane. The simple automorphic function, then, has no other singularities than poles and is, therefore, a rational function.

If the group is Fuchsian, the simple automorphic function is called a "Fuchsian function." If the group is of the first kind, the domain of existence of the function, if not constant, is the interior or the exterior of the principal circle. If the group is of the second kind, the domain of existence consists of the whole plane exclusion of the limit points lying on the principal circle.

If the group is Kleinian, the simple automorphic function is called a "Kleinian function."

Extension of the Definition.—If $f(z)$ is a simple automorphic function belonging to a group T , we shall define

$$\psi(z) = f[S(z)],$$

where S is a linear transformation to be a simple automorphic function belonging to the transformed group $S^{-1}TS$. It is clearly automorphic; for

$$\psi[S^{-1}TS(z)] = f[SS^{-1}TS(z)] = f[TS(z)] = f[S(z)] = \psi(z).$$

Here $\psi(z)$ has the domain of existence $S^{-1}(\Sigma)$, where Σ is the domain of existence of $f(z)$; and $S^{-1}(R_0)$ is a fundamental region with a finite number of sides for the transformed group.

¹ If P counts as two parabolic points of the region, as in Fig. 19, the approach shall be from one side only. There shall be a limit when the approach is from either side, but the two limits may be different.

By this means we extend the definition to groups with fixed points or limit points at infinity. Such a group can be represented in infinitely many ways as the transform of a group with infinity as a non-fixed ordinary point. The propositions derived in this chapter—on zeros, poles, algebraic relations, etc.—hold for the fundamental region $S^{-1}(R_0)$ when the group is transformed.

Also, for many of the commoner groups we know already fundamental regions, not based on isometric circles, to which the proofs of the following theorems apply (Figs. 11 to 14, for example). In the general case, however, we fall back on the region R_1 , for the reason that the properties of R have been established with complete generality.

A familiar example of a simple automorphic function is the Weierstrassian $\wp$ -function. Here the domain of existence is the finite plane, the period parallelogram is a fundamental region, and there are no parabolic points.

Likewise $\sin z$ is a simple automorphic function. The period strip is a fundamental region with parabolic points at the ends. We find readily that $\sin z$ approaches ∞ as z approaches either end of the strip.

41. Behavior at Vertices and Parabolic Points.—Of the fixed points of the transformations of a group, only those belonging to elliptic transformations can lie within the domain in which the automorphic function is analytic or has poles; all other fixed points are limit points. At a fixed point of an elliptic transformation, the function must behave in a particular way.

THEOREM 6.—*A non-constant automorphic function takes on its value k times, or some multiple thereof, at a fixed point of an elliptic transformation of period k within the domain of existence of the function.*

A function $f(z)$ is said to take on its value s times at a point z_0 at which it is analytic if it can be written, in the neighborhood of the point, in the form $f(z) = f(z_0) + (z - z_0)^s \varphi(z)$ where $\varphi(z)$ is analytic at z_0 and $\varphi(z_0) \neq 0$. It takes on the value infinity s times if $1/f(z)$ takes on the value zero s times. If $z_0 = \infty$, $z - z_0$ is replaced by $1/z$.

Let z_0 be the fixed point of an elliptic transformation of period k . We shall suppose that z_0 is finite. Then there is a transformation S of the form

$$\frac{z' - z_0}{z' - z_1} = e^{2\pi i/k} \frac{z - z_0}{z - z_1},$$

where z_1 , the second fixed point, is distinct from z_0 . Let $f(z)$ be analytic at z_0 ; then we can write, for z in the neighborhood of z_0 ,

$$f(z) = f(z_0) + (z - z_0)^s \varphi(z) = f(z_0) + \left(\frac{z - z_0}{z - z_1}\right)^s \psi(z),$$

where $\psi(z) = (z - z_1)^s \varphi(z)$; so $\psi(z)$ is analytic at z_0 and $\psi(z_0) \neq 0$. If z is in the neighborhood of z_0 so also is $z' = S(z)$; so

$$f(z') = f(z_0) + \left(\frac{z' - z_0}{z' - z_1}\right)^s \psi(z') = f(z_0) + \left(\frac{z - z_0}{z - z_1}\right)^s e^{2\pi i s/k} \psi(z').$$

Since the function is automorphic, $f(z') = f(z)$, whence

$$e^{2\pi i s/k} = \frac{\psi(z)}{\psi(z')}.$$

The first member of this equation is constant; so, also, is the second. Letting z approach z_0 , z' also approaches z_0 , and we have

$$e^{2\pi i s/k} = 1.$$

It follows that s , which is a positive integer, is a multiple of k , which was to be proved.

If $f(z)$ has a pole at z_0 , $1/f(z)$ is an automorphic function with a zero at z_0 . The order of the zero is a multiple of k ; hence, $f(z)$ has a pole whose order is a multiple of k .

The proof for the case that $z_0 = \infty$ is not essentially different.

THEOREM 7.—*A non-constant automorphic function takes on its value k times, or some multiple thereof, at a vertex belonging to a cycle the sum of whose angles is $2\pi/k$.*

This is a corollary of the preceding theorem. If $k > 1$, the vertex is a fixed point of an elliptic transformation of period k (Sec. 26, Theorem 14); if $k = 1$, the proposition is trivial.

Consider now the behaviour of a simple automorphic function at a parabolic point P . We can carry the congruent points of the cycle to P , the transforms of the fundamental region fitting together at P as in Fig. 23. Then $f(z)$ approaches a definite finite or infinite value as z approaches P from within the regions of that figure. For, $f(z)$ approaches a limit as z approaches P within or on the boundary of each region; and, owing to the common boundaries, the limits are equal. In other words, $f(z)$ approaches the same value at all the points of the parabolic cycle.

As pointed out in Sec. 27, the transformation S which carries R of Fig. 23 into R_1 , and the side l_1 into l' , is parabolic. S has the form (Sec. 10, Equation 37)

$$\frac{1}{z' - P} = \frac{1}{z - P} + c.$$

We shall investigate the function in the triangular region $z_1 z_2 P$ formed by l_1 , l' , and a small circle C through P orthogonal to the sides which meet at P (Fig. 32(a)). C is a fixed circle for S . By repeated applications of S , the transforms of the region mentioned fill up the circle C ; and the values of $f(z)$ repeat themselves in the transformed regions.

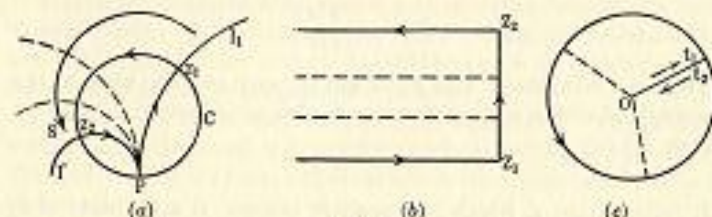


FIG. 32.

We make the change of variable

$$t = e^{2\pi i/c(z-P)},$$

thus mapping the region under consideration on the t -plane. We write this in the form

$$Z = \frac{2\pi i}{c(z-P)}, \quad t = e^Z.$$

The first transformation is linear. It carries P into ∞ and the circles into straight lines. Let Z_1 and Z_2 be the transforms of z_1 and z_2 . We have

$$Z_1 = \frac{2\pi i}{c(z_1 - P)},$$

$$Z_2 = \frac{2\pi i}{c(z_2 - P)} = \frac{2\pi i}{c} \left[\frac{1}{z_1 - P} + c \right] = Z_1 + 2\pi i.$$

The arc $z_1 z_2$ of C is transformed into a straight line parallel to the imaginary axis (Fig. 32(b)). The sides Pz_1 and $z_2 P$ are carried into straight lines perpendicular to $Z_1 Z_2$, and, hence, parallel to the real axis. The triangle of (a) is mapped on the

region bounded by the three lines of (b). Congruent points of l_1 and l' are carried into points in (b) which differ by $2\pi i$.

These latter points are carried into coincident points by the transformation $t = e^Z$. The side $Z_1 Z_2$ is carried into a circle with center at the origin. The original circular arc triangle is thus mapped on a circle slit along a radius, as in (c).

The function $f(z)$ becomes a function of t , $\varphi(t)$, analytic except for poles within and on the boundary of the region in (c) except possibly at 0. $\varphi(t)$ takes on the same value at a point on the radius along which the region is slit when approached from either side. This slit can be removed, the function being single-valued. Since $\varphi(t)$ approaches a finite value at 0 or becomes infinite, either it is analytic at 0, if properly defined there, or it has a pole at 0.

We have established the following result:

THEOREM 8.—At a parabolic point a simple automorphic function is a function of t analytic or having a pole at $t = 0$, where

$$t = e^{2\pi i/c(z-P)}.$$

If $P = \infty$, S has the form $z' = z + c$ and we put $t = e^{2\pi i/c}$ in Theorem 8.

In the neighborhood of 0, Fig. 29 (c), we can write

$$f(z) = \varphi(t) = a_0 + a_1 t + a_2 t^2 + \dots,$$

or

$$f(z) = t^{-s}(a_0 + a_1 t + \dots),$$

according as $f(z)$ approaches a finite value or becomes infinite as z approaches P . This expansion is valid in a circle with 0 as center and passing through the nearest singularity of the function in the t -plane. Carrying this back to the z -plane, the expansion is valid within the circle C through the nearest pole or limit point of $f(z)$.

42. The Poles and Zeros.—In counting the poles and zeros of a simple automorphic function which lie in the fundamental region, certain conventions are necessary in the cases of poles or zeros lying on the boundary. (1) If there is a pole or zero on the side l_n there is a pole or zero at the congruent point on the side l_{-n} . Only one of these shall be counted as belonging to the region. (2) The order s of a pole or zero at a vertex shall be partitioned equally among the regions which meet there. If there are m vertices in the cycle and the sum of the angles at the vertices is $2\pi/k$, then km regions meet at each vertex. Counting

s/k poles or zeros at each vertex, we have s/k poles or zeros at all the m vertices of the cycle. This number (Theorem 7) is an integer. (3) If $f(z)$ becomes infinite or approaches zero at a parabolic point P we shall determine the number of poles or zeros from the behavior of $\varphi(t)$ at the origin. The number of poles or zeros of $\varphi(t)$ at $t = 0$ shall be the number of poles or zeros of $f(z)$ in the parabolic points, taken all together, of the cycle to which P belongs.

The number of times $f(z)$ takes on any other value C shall be the number of times $f(z) - C$ takes on the value zero.

THEOREM 9.—A simple automorphic function which is not identically zero has an equal number of zeros and poles in the fundamental region.

Suppose, first, that the function has neither poles nor zeros on the boundary. Then,

$$N - M = \frac{1}{2\pi i} \int d \log f(z),$$

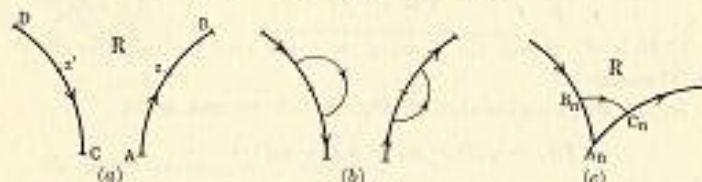


FIG. 33.

where N is the number of zeros in the fundamental region, and M the number of poles, the integral being taken in a positive sense around the boundary. (If the region is multiply connected or disconnected, we integrate, of course, around the complete boundary.)

Consider the parts of the integral arising from two congruent sides AB and CD (Fig. 33(a)): $\int_A^B d \log f(z) + \int_D^C d \log f(z)$. At the congruent points z and z' , we have $f(z') = f(z)$; then, $\log f(z')$ and $\log f(z)$ differ at most by a multiple of $2\pi i$, and $d \log f(z) = d \log f(z')$. The second integral can then be written $\int_B^A d \log f(z)$; and the two integrals cancel. The integrals along each pair of congruent sides cancel, and we have $N - M = 0$, or $N = M$; which was to be proved.

If there is a zero or pole on the side AB , we deform the side slightly, as in Fig. 33(b), so as to include the zero or pole, and

we make the corresponding alteration in the side CD . The integrals along the new congruent sides cancel as before. Only one of the two zeros or poles now lies within the contour. The theorem holds as before, since but one of the pair should be counted as belonging to the region.

Let $f(z)$ have a zero of order s at a vertex. We alter the path of integration to exclude each vertex of the cycle as in Fig. 33(c), the points B_n, C_n being at a distance d from A_n . The parts of the sides that remain are congruent in pairs; and the integrals over these sides cancel as before. At A_n we have $f(z) = (z - A_n)^s \varphi(z)$, where $\varphi(z)$ is analytic at A_n and does not vanish there; and

$$\int_{A_n}^{C_n} d \log f(z) = s \int_{A_n}^{C_n} d \log (z - A_n) + \int_{A_n}^{C_n} \frac{\varphi'(z) dz}{\varphi(z)}.$$

Letting d approach zero, the last integral approaches zero, since the integrand remains finite and the length of the path of integration approaches zero. The first integral of the second member approaches $s(-i\angle A)$. Summing for all the vertices of the cycle, we have

$$N - M = -\frac{s}{2\pi} \sum \angle A_n.$$

If the sum of the angles of the cycle is $2\pi/k$, we have

$$N - M = -\frac{s}{k}, \text{ or } N + \frac{s}{k} = M.$$

Here, N is the number of zeros within the contour. As s/k is precisely the number of zeros which we are to count at the vertices of the cycle, the number of zeros is equal to the number of poles.

A pole at a vertex is treated similarly, s being replaced by $-s$.

Finally, let $f(z)$ have a pole or zero at a parabolic point P . Draw a circle C through P , as in Fig. 32, sufficiently small that there are no poles or zeros other than at P within the region $z_1 z_2 P$ of that figure. The arc $z_1 z_2$ cuts off from the regions that lie in $z_1 z_2 P$ certain parts $A_1, A_2, A_3, \dots, A_n$. The congruent parts $A'_1, A'_2, \dots, A'_n$ lying in the fundamental region make up the neighborhoods of the parabolic points of the cycle. We shall remove these parts from the fundamental region and integrate around the contour of the remainder. The integrals over pairs of congruent sides cancel. The integrals over the circular arcs

cutting off the parabolic points may be replaced by the integrals over the congruent arcs on C :

$$N - M = \frac{1}{2\pi i} \int_{\gamma} d \log f(z).$$

This last integral may be replaced by $1/2\pi i \int d \log \varphi(t)$ taken clockwise around the circle in Fig. 32(c). This has the value $-s$ if $\varphi(t)$ has a zero of order s at $t = 0$, and the value p if $\varphi(t)$ has a pole of order p there. Hence we have $N + s = M$ or $N = M + p$; and the theorem holds.

By combinations of the preceding methods of contour integration we dispose of all cases in which there is a finite number of poles and zeros on the boundary of the fundamental region. Now, $f(z)$ cannot have an infinite number of poles for the poles would then have a cluster point z' . If z' is an ordinary point, $f(z)$ has an essential singularity there; if z' is a parabolic point, $\varphi(t)$ has an essential singularity at the origin, both of which are contrary to hypothesis. Similarly $f(z)$ cannot have an infinite number of zeros unless it is identically zero, for a cluster point of zeros would be, likewise, an essential singularity. Hence, the theorem is established.

THEOREM 10.—*A simple automorphic function which has no poles in the fundamental region is a constant.*

Let $f(z)$ be a simple automorphic function having no poles in the fundamental region; and let its value at z_1 , a point of the region, be C . Then $f(z) - C$ is a simple automorphic function with a zero at z_1 and having no poles. This is possible, according to Theorem 9, only if $f(z) - C = 0$, or $f(z) = C$.

THEOREM 11.—*A simple automorphic function which is not a constant takes on every value the same number of times in the fundamental region.*

If a simple automorphic function $f(z)$ is not constant, it has a certain finite number of poles in the fundamental region. The function $f(z) - C$, where C is any constant, is a simple automorphic function with the same poles as $f(z)$. The number of zeros is equal to the number of poles. That is, the number of times $f(z)$ takes on the value C is equal to the number of poles of $f(z)$, which establishes the theorem.

43. Algebraic Relations.—As a consequence of the preceding results, the following important theorem can now be established.

THEOREM 12.—*Between two simple automorphic functions belonging to the same group and having the same domain of existence, there exists an algebraic relation.*

Let $f_1(z)$ and $f_2(z)$ be two such simple automorphic functions, with k_1 and k_2 poles, respectively, in the fundamental region. We are to show that there exists a relation of the form

$$\Phi(f_1, f_2) = A_1 f_1^m f_2^n + A_2 f_1^{m-1} f_2^{n+1} + \dots + A_{(m+1)(n+1)} = 0, \quad (2)$$

where $A_1, A_2, \dots$ are constants, the relation holding for all values of z in the domain of definition of the functions. Whatever values be given to the constants, the function Φ is a simple automorphic function. The degrees of Φ in f_1 and f_2 , namely, m and n , respectively, will be determined later. The number of poles of Φ is not greater than $mk_1 + nk_2$.

The most general polynomial of degree m in f_1 and n in f_2 contains $(m+1)(n+1)$ constants. We can so choose these constants that Φ shall have zeros at $(m+1)(n+1) - 1$ assigned points in the fundamental region. For, let $c_1, c_2, \dots, c_{(m+1)(n+1)-1}$ be distinct points of the region different from the poles of $f_1(z)$ and $f_2(z)$; and let $A_1, A_2, \dots$ be determined to satisfy the equations

$$A_1 f_1^m(c_i) f_2^n(c_i) + \dots + A_{(m+1)(n+1)} = 0, \quad (3)$$

$$i = 1, 2, \dots, (m+1)(n+1) - 1.$$

Constants not all zero can always be found to satisfy these equations, since there is one more constant than equations to be satisfied. With these values of $A_1, A_2, \dots$, Φ has zeros at the points $c_1, c_2, \dots$.

The function then has at least $(m+1)(n+1) - 1$ zeros and not more than $mk_1 + nk_2$ poles. Now, if m and n be large enough

$$(m+1)(n+1) - 1 > mk_1 + nk_2, \quad (4)$$

and Φ has more known zeros than possible poles. According to Theorem 9, Φ is identically zero. This establishes the theorem.

It will generally happen, if the algebraic relation be found in this manner, that Φ is reducible:

$$\Phi(f_1, f_2) = \Phi_1(f_1, f_2) \Phi_2(f_1, f_2) \dots \Phi_r(f_1, f_2),$$

where Φ_i is an irreducible polynomial in f_1 and f_2 . Some one of the irreducible factors must vanish identically. This irreducible relation will contain both factors unless one of them is a constant;

for from a relation of the form $\Phi_1(f_1) = 0$, we deduce $f_1 = \text{const.}$ If neither function is a constant, there is essentially but one algebraic relation connecting them; for, from two independent relations $\Phi_1(f_1, f_2) = 0$ and $\Phi_2(f_1, f_2) = 0$, we have, on solving, $f_1 = \text{const.}$ and $f_2 = \text{const.}$

It is easy to see, in the general case, what the degrees of the irreducible equation in f_1 and f_2 will be. The degree in f_1 is the number of values of f_1 which satisfy the equation when f_2 is given a fixed value. There are k_2 points in the fundamental region at which $f_2 = C$, where, as before, k_2 is the number of poles of f_2 . At each of these points f_1 has a value satisfying the irreducible equation. Hence, in general, this equation is of degree k_2 in f_1 . Similarly, it is of degree k_1 in f_2 , in general. It may happen, for particular functions, that some of the values of f_1 at the k_2 points are always coincident, in which case the degree in f_1 is less than k_2 . The functions f_2 and $f_1 = f_2^2$ furnish a simple example.

There arises the question whether, conversely, each pair of values c_1, c_2 satisfying the irreducible equation $\Phi_1(c_1, c_2) = 0$ is taken on by the functions at some point of the fundamental region. This is, in fact, the case. The algebraic equation $\Phi_1(f_1, f_2) = 0$ determines f_1 , say, as a function of f_2 , $f_1 = \psi(f_2)$, and the Riemann surface of this function is connected. All pairs of values satisfying the equation are represented by points of the Riemann surface.

In the neighborhood of a point z_0 of the fundamental region let $f_1 = c_1', f_2 = c_2'$, the points being so chosen that $c_1'c_2'$ is not at a branch point of the Riemann surface. In the neighborhood of z_0 , $f_1(z)$ coincides with that branch of $\psi[f_2(z)]$ which takes on the value c_1' at z_0 . These two functions are then equal wherever they can be extended analytically. Now, in the f_2 -plane we can trace such a path that as f_2 moves from c_2' to c_2 , $\psi(f_2)$ moves from c_1' to c_1 . Along this path $z(f_2)$ is analytic provided we avoid certain singular positions, since the derivative $dz/df_2 = 1/f_2'(z)$ exists. Hence, z traces a path in the domain of existence of the functions. At the terminus z' of this path $f_1(z') = c_1$. In the point z_0' of the fundamental region congruent to z' we have $f_1(z_0') = c_1, f_2(z_0') = c_2$.

THEOREM 13.—Any simple automorphic function can be expressed as a rational function of two simple automorphic functions which have the property that an arbitrary pair of values of the

functions is taken on at but one point of the fundamental region, the domain of existence of the three functions being the same.

Let f_1 and f_2 be two functions with the desired property, and let f_3 be a third function. To each pair of values of f_1 and f_2 , there corresponds, in general, one value, and only one, of z in the region and, hence, one value of f_3 . That is, f_3 is single valued on the Riemann surface of $f_1 = \psi(f_2)$. Also, f_3 is an analytic function of f_2 , except for certain exceptional points where f_2 becomes infinite or $f_2'(z)$ is zero. At all exceptional points of the surface, f_3 approaches a finite value or becomes infinite. Hence, f_3 has no other singularities than poles on the Riemann surface. By a well-known theorem, f_3 is a rational function of f_1 and f_2 :

$$f_3 = \frac{A_1 f_1' f_2' + \dots}{B_1 f_1' f_2' + \dots} = R(f_1, f_2). \quad (5)$$

THEOREM 14.—If there exists a simple automorphic function $f_1(z)$ having a single pole in the fundamental region, then any simple automorphic function connected with the group, and having the same domain of definition, is a rational function of $f_1(z)$.

Let $f_2(z)$ be a simple automorphic function with k_2 poles. If in equation (2) we take $m = k_2, n = 1$, the inequality (4) is satisfied: $2k_2 + 1 > 2k_2$. The identically vanishing polynomial (2) is then of the form

$$Q_{k_2}(f_1) \cdot f_2 + P_{k_2}(f_1) = 0,$$

where P_{k_2} and Q_{k_2} are polynomials of degree k_2 at most. Not all the coefficients in Q_{k_2} are zero; for, otherwise, $P_{k_2}(f_1) = 0$ and $f_1 = \text{const.}$, contrary to hypothesis. We have, then,

$$f_2 = -\frac{P_{k_2}(f_1)}{Q_{k_2}(f_1)}, \quad (6)$$

which was to be proved.

COROLLARY.—The most general simple automorphic function having a single pole in the fundamental region is the function

$$f = \frac{Af_1 + B}{Cf_1 + D}, \quad AD - BC \neq 0.$$

It is clear that this function has a single pole in the fundamental region whatever the constants A, B, C, D may be, provided that $AD - BC \neq 0$. For, the values of f correspond in a one-to-one manner to those of f_1 ; hence, f takes on each value once in the

fundamental region. That this is the most general such function follows from the proof of the preceding theorem. Any function f_2 having a single pole satisfies an equation of the form (6) where, since $k_2 = 1$, the numerator and denominator are linear, and where, since $f_2 \neq \text{const.}$, the determinant is different from zero.

44. Differential Equations.—The derivative of an automorphic function is, in general, not automorphic. Differentiating the equation

$$f(z') = f(z),$$

where

$$z' = \frac{az + b}{cz + d}, \quad \frac{dz'}{dz} = \frac{1}{(cz + d)^2},$$

we have

$$f'(z') = f'(z) \frac{dz}{dz'} = (cz + d)^2 f'(z). \quad (7)$$

The derivative is automorphic with respect to the group, provided it is not identically zero, only if $c = 0$, $d = \pm 1$ for all transformations of the group. The transformations are then all of the form $z' = z + b$; and the group, as we shall find later (Sec. 59), is simply or doubly periodic.

We observe from (7) that the quotient of the first derivatives of two automorphic functions belonging to a group is unaltered by the transformations of the group. This is apparent for simple automorphic functions when we differentiate the relation connecting them. We have $\psi_1 f_1'(z) + \psi_2 f_2'(z) = 0$; whence the quotient $f_1'(z)/f_2'(z)$ is rational in $f_1(z)$ and $f_2(z)$, and so is a simple automorphic function.

We can set up other combinations which are unaltered by the transformations of the group. Differentiating (7), we find

$$f''(z') = (cz + d)^4 f''(z) + 2c(cz + d)^3 f'(z),$$

$$f'''(z') = (cz + d)^6 f'''(z) + 6c(cz + d)^5 f''(z) + 6c^2(cz + d)^4 f'(z),$$

and

$$2f'(z')f'''(z') - 3f''(z')^2 = (cz + d)^6 [2f'(z)f'''(z) - 3f''(z)^2]. \quad (8)$$

The quantity

$$D(y)_z = \frac{2 \frac{dy}{dz} \frac{d^2 y}{dz^2} - 3 \left(\frac{d^2 y}{dz^2} \right)^2}{2 \left(\frac{dy}{dz} \right)^3} = \frac{d^2}{dx^2} \log \frac{dy}{dx} - \frac{1}{2} \left(\frac{d}{dx} \log \frac{dy}{dx} \right)^2 \quad (9)$$

is known as the "Schwarzian derivative"¹ of y with respect to

¹ SCHWARZ, H. A., "Gesammelte Mathematische Abhandlungen," Bd. 2, p. 78. Various notations have been used for this expression: $\psi(y, x)$ by Schwarz; $\{y, x\}$ by Cayley; $\{y, x\}$ by Klein; $D(y)_z$ by Koebe.

x . We have from (8) and (7) that the function

$$\frac{2f'(z)f'''(z) - 3f''(z)^2}{2f'(z)^4} = \frac{D(f)_z}{f'(z)^3} \quad (10)$$

is unaltered by the transformations of the group.

We now show that if $f(z)$ is a simple automorphic function so also is (10). If $f(z)$ is analytic or has a pole at a point, the same is true of its derivatives and the rational combination of derivatives in (10) is analytic or has a pole at the point. The function is then analytic except for poles throughout the domain of existence of $f(z)$. There remains the question of its behavior at the parabolic points, if any occur. At such a point $f(z)$ is analytic or has a pole at the origin when expressed as a function of t , where

$$t = e^z, \quad Z = \frac{2\pi i}{c(z - P)} \quad \text{or} \quad Z = \frac{2\pi i z}{c},$$

according as the parabolic point is finite or infinite (Sec. 41).

In changing the variable, we make use of the following properties of the Schwarzian derivative, which are easily established:

$$D\left(\frac{ay + b}{cy + d}\right)_z = D(y)_z, \quad \text{whence} \quad D\left(\frac{ax + b}{cx + d}\right) = D(x)_x = 0. \quad (11)$$

$$D(y)_z = D(y)_t \left(\frac{dt}{dz} \right)^2 + D(t)_z. \quad (12)$$

Equation (11) expresses the fundamental property of the Schwarzian derivative; in fact, the derivative was originally so set up as to be invariant when y is subjected to a linear transformation. Equation (12) is the formula for the change of variable.

Making the change of variable given above and noting that

$$D(Z)_t = 0, \quad D(t)_z = -1/2,$$

we have

$$D(f)_z = D(f)_t \left(\frac{dZ}{dz} \right)^2 = \left(\frac{dZ}{dz} \right)^2 \left[D(f)_t \left(\frac{dt}{dZ} \right)^2 - \frac{1}{2} \right];$$

whence,

$$\frac{D(f)_z}{\left(\frac{df}{dz} \right)^3} = \frac{D(f)_t}{\left(\frac{df}{dt} \right)^3} - \frac{1}{2 \left(\frac{df}{dZ} \right)^3} = \frac{D(f)_t}{f'^3(t)} - \frac{1}{2 \psi'^3(t)}.$$

This function of t is analytic or has at most a pole at $t = 0$. Hence, the function (10) is a simple automorphic function.

We shall now prove the following remarkable theorem:

THEOREM 15.—If $w(z)$ is a non-constant simple automorphic function of z , then z can be expressed as a function of w by the

quotient of two solutions of a linear differential equation of the second order of the form

$$\frac{d^2\eta}{dw^2} = u\eta, \quad (13)$$

where u is an algebraic function of w , $\Phi(u, w) = 0$.

If w has a single pole of the first order in the fundamental region, u is a rational function of w .

Consider the functions

$$\eta_1 = z\sqrt{\frac{dw}{dz}}, \quad \eta_2 = \sqrt{\frac{dw}{dz}}, \quad z = \frac{\eta_1}{\eta_2}. \quad (14)$$

We shall show that

$$\frac{1}{\eta_1} \frac{d^2\eta_1}{dw^2} = \frac{1}{\eta_2} \frac{d^2\eta_2}{dw^2} = \frac{D(w)_z}{2(dw/dz)^2}. \quad (15)$$

From the relations

$$\frac{dw}{dz} = \frac{\eta_1^2}{z^2} = \eta_2^2, \quad (16)$$

we proceed to find $D(w)_z$, using the last formula of (9). Taking logarithms and differentiating, we have

$$\frac{d}{dz} \log \frac{dw}{dz} = \frac{2 \frac{d\eta_1}{dz}}{\eta_1} \frac{dw}{dz} - \frac{2}{z} = \frac{2 \frac{d\eta_2}{dz}}{\eta_2} \frac{dw}{dz}.$$

Replacing dw/dz in the second and third members by its values in (16),

$$\frac{d}{dz} \log \frac{dw}{dz} = \frac{2\eta_1 \frac{d\eta_1}{dz}}{z^2} - \frac{2}{z} = 2\eta_2 \frac{d\eta_2}{dz}. \quad (17)$$

Differentiating again and substituting as before for dw/dz ,

$$\begin{aligned} \frac{d^2}{dz^2} \log \frac{dw}{dz} &= \left[\frac{2\eta_1 \frac{d^2\eta_1}{dz^2}}{z^2} + 2 \left(\frac{d\eta_1}{dz} \right)^2 \frac{1}{z^3} \right] \eta_1^2 - \frac{4\eta_1 \frac{d\eta_1}{dz}}{z^2} + \frac{2}{z^2} \\ &= \left[2\eta_2 \frac{d^2\eta_2}{dz^2} + 2 \left(\frac{d\eta_2}{dz} \right)^2 \right] \eta_2^2. \end{aligned} \quad (18)$$

On subtracting half the square of (17) from (18), most of the terms cancel, and we have

$$D(w)_z = \frac{2\eta_1^2 \frac{d^2\eta_1}{dz^2}}{z^4} = 2\eta_2^2 \frac{d^2\eta_2}{dz^2}.$$

On dividing by $2(dw/dz)^2$ from (16) we have (15).

It follows from (15) that η_1 and η_2 , whose ratio is z , are solutions of the differential equation

$$\frac{d^2\eta}{dw^2} = u\eta,$$

where we write u for the last member of (15). But we found in (10) that u is a simple automorphic function of z . Hence, (Theorem 12) u and w are connected by an algebraic relation, $\Phi(u, w) = 0$. In particular, if w has a single pole in the fundamental region (Theorem 14), u is a rational function of w . The theorem is thus established.

CHAPTER V

THE POINCARÉ THETA SERIES

45. The Theta Series.—In the preceding chapter, we assumed the existence of automorphic functions and studied their properties. In the present chapter, we shall demonstrate their existence by the process of actually setting them up by means of series.

Let the transformations of the group be

$$z_i = T_i(z) = \frac{a_i z + b_i}{c_i z + d_i}, \quad a_i d_i - b_i c_i = 1, \quad (1)$$

$$i = 0, 1, 2, \dots,$$

the identical transformation being $z_0 = T_0(z) = z$. As an aid to simplicity in the formulae, we shall use the notation $z_{ij} = T_i(z_j) = T_i T_j(z)$, $z_{ijk} = T_i T_j T_k(z)$, etc.

We consider, first, a case whose treatment involves little difficulty, namely, the finite group. Let the group contain m transformations ($i = 0, 1, \dots, m-1$). Let $H(z)$ be any rational function of z and form the function

$$\varphi(z) = H(z) + H(z_1) + H(z_2) + \dots + H(z_{m-1}). \quad (2)$$

This function has no other singularities than poles. If we apply a transformation of the group to z , we have

$$\varphi(z_k) = H(z_k) + H(z_{1k}) + \dots + H(z_{m-1,k}).$$

Now, $z_k, z_{1k}, \dots, z_{m-1,k}$ are the set of transforms of z and, since z_k is congruent to z , this set coincides with $z, z_1, \dots, z_{m-1}$. The terms in the sum are the same as before, their order being merely interchanged; hence, $\varphi(z_k) = \varphi(z)$. The function is thus automorphic. In fact, having no other singularities than poles, it is a simple automorphic function. In like manner, any rational symmetric function of $H(z), H(z_1), \dots, H(z_{m-1})$ is a simple automorphic function.

If the group contains an infinite number of transformations and we extend (2) to an infinite number of terms, the series will

ordinarily fail to converge. For, the general term $H(z_n)$ will not even approach zero unless $H(z)$ is zero in the limit points about which the points z_n cluster. Poincaré got around the difficulty by the introduction of convergence factors. By the use of these factors we are led to a sum which is no longer automorphic but which behaves in a simple manner when a transformation of the group is applied.

Let $H(z)$ be a rational function none of whose poles is at a limit point of the group.¹ We consider the following series:

$$\theta(z) = \sum_{i=0}^{\infty} (c_i z + d_i)^{-2m} H(z_i). \quad (3)$$

This is the theta series of Poincaré.² We shall presently establish its convergence, under suitable circumstances, when m is an integer greater than 1. We shall assume its convergence and derive the basic property of the function which it defines.

If z be subjected to a transformation T_j of the group, the series becomes

$$\begin{aligned} \theta(z_j) &= \sum \left(c_j \frac{a_i z + b_i}{c_i z + d_i} + d_j \right)^{-2m} H(z_{ij}) \\ &= \sum \left[\frac{(c_j a_i + d_j c_i) z + c_j b_i + d_j d_i}{c_i z + d_i} \right]^{-2m} H(z_{ij}). \end{aligned}$$

The factor $(c_j z + d_j)^{2m}$ comes out of all the terms and we have, on replacing the numerator by an equivalent expression,

$$\theta(z_j) = (c_j z + d_j)^{2m} \sum (c_i z + d_i)^{-2m} H(z_{ij}).$$

The series on the right is the series (3) with the terms rearranged. Our subsequent convergence proof will justify rearranging the terms, so that we have

$$\theta(z_j) = (c_j z + d_j)^{2m} \theta(z). \quad (4)$$

This equation expresses the fundamental property of the theta series.

By means of these series we can set up functions which are unaltered when a transformation of the group is applied. Let

¹ In the case of a function group which carries a region Σ into itself, it suffices that $H(z)$ have no other singularities than poles in Σ and be bounded on the boundary of Σ .

² POINCARÉ, H., "Mémoire sur les fonctions fuchsienues," *Acta Math.*, vol. 1, pp. 193-294; *Oeuvres*, vol. 2, pp. 169-257.

$\theta_1(z)$ and $\theta_2(z)$ be two theta series formed with the same integer m and consider their quotient, $F(z) = \theta_1(z)/\theta_2(z)$. We have

$$F(z_i) = \frac{\theta_1(z_i)}{\theta_2(z_i)} = \frac{(c_i z + d_i)^{-2m} \theta_1(z)}{(c_i z + d_i)^{-2m} \theta_2(z)} = F(z). \quad (5)$$

It will appear subsequently that, for a function group, z_i lies in the domain of existence of the theta functions. Then $F(z)$ is an automorphic function.

Poincaré called the series (3) a "theta-fuchsian series" or a "theta-kleinian series" according as the group to which it belongs is Fuchsian or Kleinian. He calls a function with the property (4) and which has no other singularities than poles at ordinary points of the group, whatever the manner of its formation may be, a "theta-fuchsian function" or a "theta-kleinian function." For example, the derivative of an automorphic function (Sec. 44, Equation 7) is a theta function with $m = 1$.

46. The Convergence of the Series.—The following proposition is fundamental for establishing the convergence of various series and products connected with the group:

THEOREM 1.—If the point at infinity is an ordinary point of the group, the series $\sum |c_n|^{-2m}$, where in the summation the finite number of terms for which $c_n = 0$ are omitted, converges for $m > 2$.

The series in the theorem can be written $\sum r_n^{-2m}$, where r_n is the radius of the isometric circle I_n of T_n . Suppose, first, that infinity is not a fixed point for an elliptic transformation. Then, there are no points congruent to infinity in the neighborhood of infinity, and we have the groups for which we developed the properties of the isometric circles in Chap. II.

It will suffice to prove the theorem for $m = 2$. Except, possibly, for a finite number of terms $r_n < 1$, so that if $m > 2$, $r_n^{-2m} < r_n^{-4}$ and the convergence follows from that of $\sum r_n^{-4}$.

The centers of the isometric circles lie in a finite region and their radii are bounded (Sec. 17). Hence there exists a positive constant k such that a circle Q_n of radius k concentric with any isometric circle I_n contains all the isometric circles on its interior. Let Q_n' be the transform of Q_n by T_n . Q_n' is got by inverting Q_n in I_n and making certain other transformations which do not alter magnitudes (Theorem 19, Sec. 11). By inversion, we find the radius of Q_n' to be r_n^3/k ; its area is $\pi r_n^4/k^2$. Now, the exterior of Q_n' , which lies entirely within R , goes into the interior

of Q_n' , which, therefore, lies within R_n , the transform of R by T_n . As the regions R_n fit together without overlapping, it follows that there is no overlapping of the circles Q_n' . Their areas then are less than the area of any one of the circles Q_n , which encloses them. We have then

$$\sum \frac{\pi r_n^4}{k^2} < \pi h^2, \quad \sum r_n^4 < h^4,$$

and the series converges.

If infinity is a fixed point for an elliptic transformation, the reasoning is not essentially different. The isometric circles are confined to a finite region as before, and the constant k exists. There is the difference that a point outside Q_n may have $p - 1$ points exterior to Q_n congruent to it, where the elliptic cyclic subgroup with fixed point at ∞ contains p transformations. The circles Q_n' can overlap, but no point can be interior to more than p such circles. Hence,

$$\sum \frac{\pi r_n^4}{k^2} < p\pi h^2, \quad \sum r_n^4 < ph^4,$$

and the series is convergent.

By the use of the preceding theorem, we are able to establish the convergence of the theta series.

THEOREM 2.—If $m > 2$ and if the point at infinity is an ordinary point of the group, then the theta series (3) defines a function which is analytic except possibly for poles in any connected region not containing limit points of the group in its interior.

It will suffice to prove the theorem for a region S' such that there are no limit points of the group within or on the boundary of S' , since such a region can be made large enough to include any given interior point of a region with limit points on its boundary.

We observe, first, that certain terms of (3) may have poles in S' . At $z = -d_i/c_i$, the center of the isometric circle I_i , the factor $(c_i z + d_i)^{-2m}$ becomes infinite. Again, if z is such that $z_i = a$, where a is the pole of $H(z)$,—that is, if $z = T_i^{-1}(a)$ —then $H(z_i)$ has a pole. It is clear, however, that only a finite number of terms of the series have poles in S' ; for S' contains in its interior and on its boundary only a finite number of centers of isometric circles and only a finite number of points congruent to each of the poles of $H(z)$.

We now put aside the finite number of terms having poles in and on the boundary of S' , and we prove that the remainder of

the series converges absolutely and uniformly in S' . Let $d > 0$ be the minimum distance from the boundary of S' to the centers of those isometric circles whose centers are exterior to S' . We have, then, for all the terms we are considering and for all z in S'

$$\left| z + \frac{d_i}{c_i} \right| \geq d.$$

Further, we can draw curves about the poles of $H(z)$ such that when z is in S' all points z_i in the terms considered, lie outside the several curves. But in the regions outside these curves, $H(z)$ is bounded, so that we have

$$|H(z_i)| < M.$$

We have, then, excluding further the finite number of terms for which $c_i = 0$,

$$|(c_i z + d_i)^{-2m} H(z_i)| = \left| \frac{H(z_i)}{c_i^{2m} \left(z + \frac{d_i}{c_i} \right)^{2m}} \right| < \frac{M}{d^{2m}} |c_i|^{-2m}. \quad (6)$$

This inequality holds for all points of S' . Since the series of positive constant terms $\sum_{d^{2m}} |c_i|^{-2m}$ converges (Theorem 1), the

absolute and uniform convergence of the series in S' is established. The sum of the series is an analytic function in S' . It follows that (3) is analytic in S' except for the finite number of poles which arise from the terms which we put aside.

47. The Convergence for the Fuchsian Group of the Second Kind.—It can be shown that in general the theta series does not converge if $m = 1$. There are certain groups, however, for which the series does converge. An important case is the following:¹

THEOREM 3.—*For a Fuchsian group of the second kind for which the point at infinity is an ordinary point, the series $\Sigma |c_n|^{-2m}$ converges.*

Then, the theta series (3) converges for $m = 1$.

We shall suppose, first, that the principal circle is a straight line—the real axis, for example. The region R contains a portion of the real axis in the neighborhood of infinity in its interior. The isometric circles lie in a finite region, and their centers are on the real axis.

We shall employ the constructions used in the proof of Theorem 1. The circle Q_n , concentric with I_n and of suitable radius h

¹ BURNSIDE, W. "On a Class of Automorphic Functions," *Proc. London Math. Soc.*, vol. 23, pp. 49–88, 1892.

lies in R . Its exterior is transformed into the interior of the circle Q_n' , with center on the real axis. The portion of the real axis exterior to Q_n is transformed into the portion of the real axis within Q_n' . The length of this latter segment is $2r_n^2/h$, the diameter of Q_n' . Since the circles Q_n' are exterior to one another and all lie in any one of the circles Q_n , the segments of the real axis which they contain are non-overlapping, and the sum of their lengths is finite:

$$\sum_h \frac{2r_n^2}{h} < 2h, \quad \sum r_n^2 < h^2.$$

Hence, the series Σr_n^2 , or $\Sigma |c_n|^{-2}$, converges.

The proof for the general Fuchsian group of the second kind rests on the following lemma:

LEMMA.—*If the point at infinity is an ordinary point for a group T_n and its transform $T_n' = GT_n G^{-1}$, then the series $\Sigma |c_n|^{-2m}$ and $\Sigma |c_n'|^{-2m}$ both converge or both diverge.*

Let the transformation which is applied to the group be $G = (\alpha z + \beta)/(\gamma z + \delta)$, $\alpha\delta - \beta\gamma = 1$. Then (Equation (2), Sec. 15)

$$c_n' = -\gamma\delta a_n + \gamma^2 b_n - \delta^2 c_n + \gamma\delta d_n. \quad (7)$$

Suppose, first, that $\gamma \neq 0$. Substituting for b_n the value $b_n = (a_n d_n - 1)/c_n$ and combining terms, we can write (7) in the form

$$c_n' = \gamma^2 c_n \left[\left(\frac{d_n}{c_n} - \frac{\delta}{\gamma} \right) \left(\frac{a_n}{c_n} + \frac{\delta}{\gamma} \right) - \frac{1}{c_n^2} \right]. \quad (8)$$

Now, the centers, $-d_n/c_n$ and a_n/c_n , of the isometric circles lie in a finite region; and their distances from the point $-\delta/\gamma$ are bounded

$$\left| \frac{d_n}{c_n} - \frac{\delta}{\gamma} \right| < K, \quad \left| \frac{a_n}{c_n} + \frac{\delta}{\gamma} \right| < K.$$

Also, $1/|c_n|^2$ is bounded: $1/|c_n|^2 < K'$. Hence, we have

$$|c_n'| < |\gamma|^2 \cdot |c_n| (K^2 + K') = K_1 |c_n|,$$

and

$$|c_n|^{-2m} < K_1^{2m} |c_n'|^{-2m}.$$

It follows from this inequality that if $\Sigma |c_n'|^{-2m}$ converges, so also does $\Sigma |c_n|^{-2m}$.

If $\gamma = 0$, we have at once $|c_n'| = |\delta|^2 \cdot |c_n|$, δ being different from zero, and the conclusion follows as before.

Now the group T_n is also a transform of T_n' ; thus $T_n = G^{-1} T_n' G$. Hence, if $\Sigma |c_n|^{-2m}$ converges, so, also, does $\Sigma |c_n'|^{-2m}$. This establishes the lemma.

We return now to the general Fuchsian group of the second kind. Let the group be transformed so that the principal circle is carried into the real axis and an ordinary point on the principal circle is carried to infinity. We found above that $\sum |c_n|^{-2}$ converges for the transformed group. Hence, applying the lemma, $\sum |c_n|^{-2}$ converges for the original group.

The convergence of the theta series follows from Equation (6) with $m = 1$ since the series of positive terms $\sum \frac{M}{d^2} |c_n|^{-2}$ converges.

48. Some Properties of the Theta Functions.—Let Σ be a connected region of the z -plane bounded by limit points. Then, $\theta(z)$ in (3) is analytic except for poles in Σ . If the limit points separate the plane into two or more regions, the series (3) defines a function in each region, but, in general, the functions so defined are distinct.

Consider, for example, the Fuchsian group of the first kind. Here the limit points consist of all points of the principal circle. Then (3) defines a function analytic except for poles within the principal circle. The poles of $\theta(z)$ arise from the poles of the individual terms of the series. If $H(z)$ has a pole at a within the principal circle, then $H(z_i)$ becomes infinite when $z_i = a$. That is, $\theta(z)$ has poles at the poles of $H(z)$ within the principal circle and at points congruent to these poles—except that, in special cases, $H(z)$ may have poles at congruent points of such a character that the singularities arising from two or more terms of the series cancel. Putting this special case aside, the number of poles in the fundamental region R_0 is exactly equal to the number of poles of $H(z)$ within the principal circle. If $\theta(z)$ has poles within the principal circle, these poles cluster in infinite numbers about each point on the principal circle. The principal circle is thus a natural boundary of the function.

In a similar manner, (3) defines a function analytic except for poles on the exterior of the principal circle. These poles are the poles of $H(z)$ lying without the principal circle, the points congruent thereto, and the points $-d_i/c_i$, where a convergence factor becomes infinite; although here, also, there is a possibility of the singularities from different terms cancelling. If the function has poles, these poles cluster about the points of the principal circle, and the function cannot be extended analytically across the circle. The one formula (3) then defines two distinct

functions, one existing within the principal circle, the other without.

In the Fuchsian group of the second kind the function defined by (3) is a single function, existing both within and without the principal circle.

The proof of the convergence of the theta series does not require that the group be a function group. The property (4) is a relation connecting two distinct functions, however, unless $\theta(z)$ can be continued analytically from z to z_j . In order that (4) express a property of a single one of the functions defined by the series, it is necessary that the domain of existence of the function be carried into itself by the transformations of the group. The group is then a function group.

In setting up functions by means of the series (3), the poles of $H(z)$ are at our disposal. By placing a pole at a desired point, we can be sure that $\theta(z)$ in a region Σ under consideration has a singularity and, hence, is not identically zero. Further, in forming automorphic functions for a function group by means of (5), we can place the poles of the numerator and denominator at different points of Σ and, thus, be assured that the automorphic function does not reduce to the trivial case of a constant.

Equation (4) exhibits a relation between the theta function and the isometric circle. If $\theta(z)$ is neither zero nor infinite at z , then $|\theta(z_j)|$ is greater than, less than, or equal to $|\theta(z)|$ according as $|cz + d|$ is greater than, less than, or equal to 1; that is, according as z lies without, within, or on the isometric circle of T . We note that $|\theta(z)|$, when not zero or infinity, has a smaller value at a point of R (exterior to all isometric circles) than at any congruent point within the domain of existence of the function.

Behavior at a Vertex.—Let ξ be the fixedpoint of an elliptic transformation lying within the domain of existence of a theta function. The transformations with ξ as fixed point form an elliptic cyclic group generated by $z' = T(z)$ of the form

$$\frac{z' - \xi}{z' - \xi'} = K \frac{z - \xi}{z - \xi'}, \quad K = e^{2\pi i/k}, \quad (9)$$

where ξ' is the second fixed point and k is an integer greater than 1. The fundamental property (4), for T , may be written

$$\theta(z') = \left(\frac{dT}{dz} \right)^{-n} \theta(z). \quad (10)$$

We find, on taking the logarithm of (9) and differentiating,

$$\frac{dT}{dz} = \frac{dz'}{dz} = \frac{(z' - \xi)(z' - \xi')}{(z - \xi)(z - \xi')}. \quad (11)$$

If we put

$$P(z) = (z - \xi)^m (z - \xi')^{-m} \theta(z), \quad (12)$$

we can write (10) in the form

$$\begin{aligned} P(z') &= (z' - \xi)^m (z' - \xi')^{-m} \theta(z') \\ &= (z - \xi)^m (z - \xi')^{-m} \theta(z) \\ &= P(z). \end{aligned} \quad (13)$$

Hence, $P(z)$ is unaltered by the transformation T , and consequently is automorphic with respect to the cyclic subgroup generated by T .

Applying Theorem 6, Sec. 41, we conclude that $P(z)$ takes on its value k times, or a multiple thereof, at ξ .

Suppose that $\theta(z)$ is analytic at ξ . Then $F(z)$ has a zero at ξ of order m at least, owing to the factor $(z - \xi)^m$. Now, if m is not a multiple of k , then $\theta(z)$ must have a zero at ξ also, the order s of the zero being such that

$$m + s = lk, \quad (14)$$

where l is an integer.

If $\theta(z)$ is analytic at ξ , it is necessarily zero there unless m is a multiple of k . The order s of the zero satisfies an equation of the form (14).

If $\theta(z)$ has a pole of order p at ξ , then $P(z)$ has a pole of order $p - m$ or a zero of order $m - p$, unless $p = m$. An equation of the form (14) holds where we put $s = -p$, l being a positive or negative integer or zero.

Behavior at a Parabolic Point.—There is a parabolic cyclic subgroup generated by a transformation S with the parabolic point P as fixed point. S has the form

$$\frac{1}{z' - P} = \frac{1}{z - P} + c. \quad (15)$$

Here we have

$$\frac{dS}{dz} = \frac{dz'}{dz} = \left(\frac{z' - P}{z - P} \right)^2. \quad (16)$$

Writing

$$G(z) = (z - P)^{2n} \theta(z), \quad (17)$$

the fundamental relation (10) for the transformation S may be written in the form

$$G(z') = (z' - P)^{2n} \theta(z') = (z - P)^{2n} \theta(z) = G(z). \quad (18)$$

The function $G(z)$ is automorphic with respect to the subgroup generated by S .

On making the customary change of variable

$$t = e^{2\pi i c(z-P)} \quad (19)$$

(See Sec. 41; here we use particularly Fig. 32), $G(z)$ is transformed into a function of t , $G(z) = g(t)$, single valued in the neighborhood of $t = 0$. We have, then,

$$\theta(z) = \frac{G(z)}{(z - P)^{2n}} = \left(\frac{c \log t}{2\pi i} \right)^{2n} g(t). \quad (20)$$

The theta function thus has a logarithmic singularity at $t = 0$.

We consider now the form of $g(t)$ when $\theta(z)$ is defined by the theta series (3). We can take the circle C of Fig. 32(a) small enough that it contains no point $-d_i/c_i$ and no point congruent to a pole of $H(z)$. For C , when small enough, contains only points congruent to points of the fundamental region R which lie in the neighborhoods of the parabolic points of the cycle; and these neighborhoods can be sufficiently restricted to exclude ∞ and the finite number of points of R congruent to poles of $H(z)$. Taking C slightly smaller we have bounds for the following quantities:

$$\left| \frac{z - P}{z + \frac{d_i}{c_i}} \right| < K, \quad |H(z_i)| < M,$$

where z lies within or on the boundary of the triangle $z_1 z_2 P$ of Fig. 32(a), exclusive of P itself. In the series

$$G(z) = \sum \left(\frac{z - P}{z + \frac{d_i}{c_i}} \right)^{2n} \frac{1}{c_i^{2n}} H(z_i),$$

the general term is less in absolute value than the corresponding term of the convergent series of positive terms $K^{2n} M \sum |c_i|^{-2n}$. $G(z)$ thus remains finite as z approaches P from the interior of the triangle. In the t -plane, then, $g(t)$ is analytic in the neighborhood of $t = 0$ and is bounded; hence, $g(t)$ is analytic at $t = 0$ if properly defined there.

If we form an automorphic function as the quotient of two theta series, which may be written

$$f(z) = \frac{(z - P)^{2n} \theta_1(z)}{(z - P)^{2n} \theta_2(z)} = \frac{g_1(t)}{g_2(t)},$$

we observe that $f(z)$, as a function of t , is analytic or has a pole at $t = 0$. Then as z approaches P from the interior of R $f(z)$ approaches a definite finite or infinite limit. It thus satisfies the requirements laid down in Sec. 40 for the behavior of a simple automorphic function at a parabolic point.

49. Zeros and Poles of the Theta Functions.—We consider now function groups in which the region R_0 lying in Σ has a finite number of sides (Sec. 40). We consider theta functions having a finite number of poles in R_0 , such as those defined by the theta series (3). We suppose that $g(t)$ in (20), as in the case of the series (3), is analytic at $t = 0$ or has a pole there; and we assume further that $\theta(z)$ is not identically zero.

We make the same conventions about counting zeros and poles at ordinary points on the boundary as were made in Sec. 42. We observe that this may lead to the count of a fractional number of zeros or poles in the region; thus, at the vertices of a cycle of angle $2\pi/k$ the zeros, from (14), amount to

$$\frac{s}{k} = l - \frac{m}{k},$$

which may be fractional. The order of the zero or pole of $g(t)$ at $t = 0$ will be counted as the number of zeros or poles of the function $\theta(z)$ in the points of the parabolic cycle to which P belongs.

We now remove the neighborhoods of the vertices and parabolic points of R_0 by small circles as explained in Sec. 42. Let N_0, M_0 be the number of zeros and poles, respectively, in the resulting region. Then,

$$N_0 - M_0 = \frac{1}{2\pi i} \int d \log \theta(z), \quad (21)$$

the integral being taken around the boundary—further altered as in Fig. 33(b) if a pole or zero lies on a side.

Before evaluating the integrals along the sides, we shall consider the integrals along the small circular arcs just constructed. When we let the radii of the arcs cutting off the vertices approach zero, we find, exactly as in Sec. 42, that the integrals along these arcs approach $m_0 - n_0$ where n_0 and m_0 are the number of zeros and poles, respectively, that we assign to the various vertices.

The parabolic points, as usual, are less simple to handle. The circular arcs $k, k_1, \dots, k_n$ about the parabolic points

of a cycle $P, P_1, \dots, P_n$ are the transforms of $k, k_1, \dots, k_n$ which fit together to make up the arc $z_1 z_2$ of Fig. 32(a). The arc k_n is the transform of k_n' by some transformation $\eta_n = T_n(z)$, and we have,

$$\begin{aligned} \int_{k_n} d \log \theta(\eta_n) &= \int_{k_n'} d \log (c_k z + d_k)^{2m} \theta(z) \\ &= \int_{k_n'} d \log \left(\frac{c_k z + d_k}{z - P} \right)^{2m} G(z) \\ &= 2m \int_{k_n'} \left[d \log \left(z + \frac{d_k}{c_k} \right) - d \log (z - P) \right] + \int_{k_n'} d \log g(t), \quad (22) \end{aligned}$$

where k_n'' is the corresponding arc on the circle in the t -plane in Fig. 32(c).

We need consider here only the imaginary part of the integral $\int d \log \theta$, since the members of (21) are necessarily real. Now, when we let the radius of C in Fig. 32(a) approach zero, it is easy to see that the imaginary parts of

$$\int_{k_n'} d \log \left(z + \frac{d_k}{c_k} \right), \quad \int_{k_n'} d \log (z - P)$$

both approach zero, being i times the angle through which the line from $-d_k/c_k$, or from P to z , turns as z moves along the arc k_n' lying on $z_1 z_2$.

We have, then, only the last integral in (22) to consider. Adding the integrals arising from the several vertices, we have

$$\frac{1}{2\pi i} \int d \log g(t),$$

the integral being taken around the circle in Fig. 32(c) in a clockwise direction. Its value is $-n'$ or m' , where $g(t)$ has a zero of order n' or a pole of order m' at $t = 0$.

On shifting the integrals evaluated to the first member of (21), and letting N, M represent the total number of zeros and poles, respectively, belonging to R_0 , we have

$$N - M = \frac{1}{2\pi i} \sum \int d \log \theta(z), \quad (23)$$

the integrals being taken along the sides l_i of R_0 .

Consider the two integrals arising from congruent sides AB, CD (Fig. 33 (a) or (b)). Let $z' = (az + b)/(cz + d)$ be

the transformation carrying AB to CD . Here, $-d/c$ is the center of the arc AB . We have

$$\int_{AB} d \log \theta(z) + \int_{DC} d \log \theta(z') = \int_{AB} d \log \theta(z) + \int_{BA} d \log (cz + d)^{2m} \theta(z) = 2m \int_{BA} d \log \left(z + \frac{d}{c} \right) = 2i\alpha,$$

where α is the angle subtended by the arc BA at its center. The equal arc CD also subtends the angle α at its center; and we can write the integral as m times the sum of the angles subtended by the two arcs. We get similar results for the other sides, which gives the following theorem:

THEOREM 4.—Let N, M be the number of zeros and poles, respectively, of the theta function in the region R_0 ; and let $\alpha_1, \alpha_2, \dots, \alpha_m$ be the angles subtended by the sides of R_0 at the centers of the isometric circles on which these sides lie. Then

$$N - M = \frac{m}{2\pi} \sum \alpha_i. \quad (24)$$

We observe that the function has always more zeros than poles in R_0 . We note that the difference $N - M$ depends only upon m and upon the character of the group. It is independent of the function $H(z)$ used in the construction of the theta series (3). This independence could have been easily foreseen. The fact that the automorphic function in (5) has the same number of zeros as poles requires that the difference $N - M$ for the functions $\theta_1(z)$ and $\theta_2(z)$ be the same.

The poles and zeros in a region R_i congruent to R_0 are determined from (4). If R_0 does not contain the point ∞ the number of zeros and poles in R_i is the same as in R_0 . If, however, R_0 contains ∞ , $N - M$ is decreased by $2m$, owing to the factor $(cz + d)^{2m}$ appearing in (4).

Suppose R_0 is a fundamental region with n pairs of sides for the Fuchsian group of the first kind. Let $2\pi/k_1, 2\pi/k_2, \dots$ be the sum of the angles in the ordinary cycles of R_0 . Then the sum of the interior angles in the circular polygon is $2\pi \sum \frac{1}{k_i}$.

Consider the rectilinear polygon of $4n$ sides in Fig. 34. The sum of its interior angles is $2\pi(2n - 1)$. The sum of the angles at the centers of isometric circles is $\sum \alpha_i$; the sum at the remaining vertices is $2\pi \sum \frac{1}{k_i} + 2\pi$. Hence,

$$2\pi(2n - 1) = \sum \alpha_i + 2\pi \sum \frac{1}{k_i} + 2\pi.$$

Substituting the value of $\sum \alpha_i$ from this into (24), we have

$$N - M = m \left(n - 1 - \sum \frac{1}{k_i} \right).$$

This result is given by Poincaré in the memoir previously cited.

A particularly simple application of (24) is to the group of Sec. 25(a). Here, R is bounded by $2n$ complete circles, and $\alpha_i = 2\pi$. For R , then,

$$N - M = 2\pi n.$$

Here, R contains ∞ ; so for R_i , any congruent region,

$$N - M = 2\pi(n - 1).$$

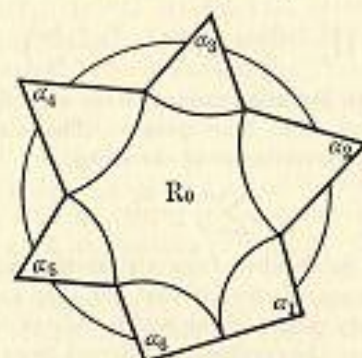


FIG. 34.

50. Series and Products Connected with the Group.—By the aid of Theorems 1 and 3 we can establish the convergence of numerous series and products connected with the group. We illustrate with a few examples.

Let $z_n, u_n, v_n, \dots$ be the transforms of $z, u, v, \dots$ by the transformation T_n of the group. We have

$$z_n - u_n = \frac{z - u}{(cz + d)(cu + d)} = \frac{z - u}{\left(z + \frac{d}{c}\right)\left(u + \frac{d}{c}\right)} \cdot \frac{1}{c^2}. \quad (25)$$

If z and u be restricted to lie in regions containing no limit points of the group and having no limit points on the boundary, then, excepting for a finite number of values of n , the factor preceding $1/c_n^2$ in (25) is bounded.

Suppose that $\sum |c_n|^{-2}$ converges; and consider the series

$$\sum (z_n - u_n),$$

the summation being extended to all transformations of the group. Except for a finite number of terms, this series converges

absolutely and uniformly in the regions mentioned. It is thus analytic in each variable except possibly for poles arising from a finite number of individual terms. In a similar manner we establish the convergence of such functions as

$$\sum \frac{z_n - u_n}{z - w_n}, \quad \sum (z_n - u_n)H(z_n),$$

where $H(z)$ is defined as for the theta series; etc.

In a subsequent chapter (Sec. 99), we shall employ products of the type

$$\prod \frac{z - u_n}{z - v_n} = \prod \left(1 + \frac{v_n - u_n}{z - v_n} \right) \quad (26)$$

in connection with Fuchsian groups of the second kind, where u and v are distinct from limit points. The convergence here depends upon the convergence of the series

$$\sum \frac{v_n - u_n}{z - v_n},$$

which is readily established from the known fact that $\sum |c_n|^{-2}$ converges. We conclude readily that (26) is an analytic function of z at all ordinary points of the z -plane except for poles at the points v_n , and that the function is different from zero except at the set of points u_n .

In the general case, where $\sum |c_n|^{-4}$ converges, we establish readily the convergence of such series as

$$\sum (z_n - u_n)^2, \quad \sum \frac{(z_n - u_n)(v_n - w_n)}{z - t_n}, \quad \sum \frac{z_n - u_n}{(c_n z + d_n)^2} H(z_n),$$

and the like.

From (26) we can build a convergent product for the general case by the insertion of convergence factors, thus

$$\prod \frac{z - u_n}{z - v_n} e^{\frac{u_n - v_n}{z - v_n}},$$

where w is an ordinary point. Another example in which the convergence is easily established is the product due to Whittaker¹

$$(z - u) \prod \left[\frac{(z - u_n)(u - z_n)}{(z - z_n)(u - u_n)} \right]^{\frac{1}{2}} e^{\frac{1}{2} \frac{(z - u)(z_n - u_n)}{(z - z_n)(u - u_n)}}$$

where the product extends to all transformations of the group excepting the identical transformation.

¹ *Mem. Math.*, vol. 31, p. 145, 1902.

CHAPTER VI

THE ELEMENTARY GROUPS

I. THE FINITE GROUPS

51. Inversion in a Sphere.—The study of finite groups will be much simplified by the introduction of transformations of three dimensional space. In this treatment, the space transformation known as "inversion in a sphere" plays a fundamental rôle. We shall now make a digression and derive the salient properties of this transformation.

Inversion in a sphere is a direct generalization of inversion in a circle (Sec. 5). Consider a sphere S with center K and radius ρ . Let P be any point in space and construct the half line beginning at K and passing through P . Let P_1 be a point on this half line such that $KP_1 \cdot KP = \rho^2$; then P_1 is called the *inverse of P with respect to S* . P is, also, obviously the inverse of P_1 . S and K are called the "sphere of inversion" and the "center of inversion," respectively. The inversion carries a point within S into a point without S and leaves the points on the surface of S fixed.

It is clear also, as in Sec. 5, that any sphere or circle through P and P_1 is orthogonal to S . For, any tangent KT' from K to the sphere or circle (Fig. 1) has the property that $\overline{KT}^2 = KP_1 \cdot KP = \rho^2$. That is, T lies on S ; so that at T the sphere or circle is orthogonal to S . Further, any other secant through K , meeting the sphere or circle at P_1' , P' has the property that $KP_1' \cdot KP' = \overline{KT}^2 = \rho^2$. Hence, by the inversion, any circle or sphere through two inverse points is transformed into itself.

We now derive the analytic expressions for the inversion. Let K, P, P_1 be the points (a, b, c) , (ξ, η, ζ) , (ξ', η', ζ') , respectively, in rectangular space coordinates. Put

$$\begin{aligned} r^2 &= (\xi - a)^2 + (\eta - b)^2 + (\zeta - c)^2, \\ r'^2 &= (\xi' - a)^2 + (\eta' - b)^2 + (\zeta' - c)^2. \end{aligned} \quad (1)$$

Then it is required that

$$rr' = \rho^2. \quad (2)$$

Projecting the segments KP and KP_1 on the three axes, we have, from the resulting similar triangles,

$$\frac{\xi' - a}{\xi - a} = \frac{\eta' - b}{\eta - b} = \frac{\zeta' - c}{\zeta - c} = \frac{r'}{r} = \frac{r'}{r^2} = \frac{\rho^2}{r^2}. \quad (3)$$

These give the following equations for the transformations

$$\xi' - a = \frac{\rho^2(\xi - a)}{r^2}, \quad \eta' - b = \frac{\rho^2(\eta - b)}{r^2}, \quad \zeta' - c = \frac{\rho^2(\zeta - c)}{r^2}. \quad (4)$$

Expressions for ξ, η, ζ in terms of ξ', η', ζ' are got by interchanging the primed and unprimed variables in (4).

Consider, now, the magnification of an element ds . Let ds' be the length of the transformed element. We have from (4)

$$d\xi' = \frac{\rho^2}{r^2} d\xi - \frac{2\rho^2(\xi - a)}{r^3} dr,$$

with similar expressions for $d\eta'$ and $d\zeta'$. Squaring and adding,

$$\begin{aligned} ds'^2 &= d\xi'^2 + d\eta'^2 + d\zeta'^2 \\ &= \frac{\rho^4}{r^4} (d\xi^2 + d\eta^2 + d\zeta^2) - \frac{4\rho^4}{r^4} dr[(\xi - a)d\xi + (\eta - b)d\eta + (\zeta - c)d\zeta] \\ &\quad + \frac{4\rho^4}{r^4} dr^2[(\xi - a)^2 + (\eta - b)^2 + (\zeta - c)^2] \\ &= \frac{\rho^4}{r^4} ds^2 - \frac{4\rho^4}{r^4} dr[rdr] + \frac{4\rho^4}{r^4} dr^2[r^2] = \frac{\rho^4}{r^4} ds^2, \end{aligned}$$

or

$$ds' = \frac{\rho^2}{r^2} ds. \quad (5)$$

The magnification depends only upon the position of the element and not upon its direction. A small triangle in the neighborhood of a point will have all its sides multiplied by the same quantity; hence, it is transformed into a similar triangle. Its angles are unaltered in magnitude. In other words, *inversion in a sphere is a conformal transformation*.

We consider now the transform of a sphere or plane Σ :

$$A(\xi^2 + \eta^2 + \zeta^2) + B\xi + C\eta + D\zeta + E = 0.$$

This is a sphere (possibly of imaginary radius) if $A \neq 0$; a plane if $A = 0$. This can be written equally well—thinking of the origin as translated to (a, b, c) —in the form

$$A[(\xi - a)^2 + (\eta - b)^2 + (\zeta - c)^2] + B'(\xi - a) + C'(\eta - b) + D'(\zeta - c) + E' = 0,$$

or

$$Ar^2 + B'(\xi - a) + C'(\eta - b) + D'(\zeta - c) + E' = 0. \quad (6)$$

Then, from (2) and (4), the transform is

$$A \frac{\rho^4}{r^4} + B' \frac{\rho^2}{r^2} (\xi - a) + C' \frac{\rho^2}{r^2} (\eta - b) + D' \frac{\rho^2}{r^2} (\zeta - c) + E' = 0,$$

or

$$E'r^2 + B'\rho^2(\xi - a) + C'\rho^2(\eta - b) + D'\rho^2(\zeta - c) + A\rho^4 = 0. \quad (7)$$

This is the equation of a sphere or plane according as $E' \neq 0$ or $E' = 0$. We have proved that *by inversion in a sphere a sphere or plane is carried into a sphere or plane*.

In (6) $E' = 0$, if Σ passes through the center of inversion. Hence, a sphere or plane through the center of inversion is transformed into a plane. This could have been foreseen geometrically.

Since a circle or straight line is the intersection of two spheres or two planes, and these latter are carried into spheres or planes, it follows that *by an inversion in a sphere a circle or straight line is carried into a circle or straight line*. The transform is a straight line if, and only if, the original circle or straight line passes through the center of inversion.

Finally, if P and Q are two points inverse with respect to a sphere Σ , the transformed points P_1 and Q_1 are inverse with respect to the transformed sphere Σ_1 . This is most easily established geometrically. Through P and Q pass a family of spheres orthogonal to Σ . The transforms of these spheres pass through P_1 and Q_1 and are orthogonal to Σ_1 —since angles are preserved. On inverting in Σ_1 , each of these spheres is transformed into itself. The point P_1 , common to all these spheres, is transformed into the second common point Q_1 . Hence, P_1 and Q_1 are inverse with respect to Σ_1 .

If Σ_1 is a plane, two points P_1 and Q_1 such that all spheres through them are orthogonal to the plane must be equidistant from the plane and on a common perpendicular to it. That is, Q_1 is the reflection of P_1 in Σ_1 . We thus extend the idea of inversion to the case in which the sphere of inversion is a plane. *Inversion in a plane is merely a reflection in a plane*. This transformation possesses the various properties of inversion in a sphere which are italicized in this section.

52. Stereographic Projection.—We now introduce a widely used method of representing the complex z -plane on a sphere. Let the complex plane be a plane in three dimensional space, and

let an inversion be made in a sphere whose center K does not lie in the z -plane. Then the z -plane is carried into a sphere which passes through K . The points of the z -plane and the points of the resulting sphere correspond in a one-to-one manner. Corresponding points lie on a line through K . Corresponding angles are equal. This correspondence between the points of the plane and of the sphere is known as a *stereographic projection*.

For convenience we shall so choose the sphere of inversion that the transform of the z -plane is the unit sphere with center at the origin. Let the space axes be so placed that the ξ - and η -axes coincide with the real and imaginary axes in the z -plane. Let K be the point $(0, 0, 1)$; and let the radius of S , the sphere of inversion, be $\sqrt{2}$. Then S passes through the unit circle Q_0 with center at the origin in the z -plane. The points of Q_0 remain fixed on making the inversion. The point at infinity in the z -plane inverts into K , the center of S . The z -plane then inverts into a sphere Σ_0 through Q_0 and the point $(0, 0, 1)$. Hence, Σ_0 is the unit sphere

$$\xi^2 + \eta^2 + \zeta^2 = 1. \quad (8)$$

Corresponding points of the z -plane and of Σ_0 lie on a line through $(0, 0, 1)$. The interior of Q_0 is transformed into the lower half of Σ_0 ; the exterior into the upper half. Circles and straight lines in the z -plane are transformed into circles on Σ_0 .

The equations connecting a point $z = x + iy$ and the corresponding point (ξ, η, ζ) of Σ_0 can be written down at once from (4). Putting $(x, y, 0)$ for (ξ, η, ζ) and (ξ, η, ζ) for (ξ', η', ζ') we have, since

$$r^2 = z^2 + y^2 + 1 = z\bar{z} + 1$$

and $\rho^2 = 2$, the equations

$$\xi = \frac{2x}{z\bar{z} + 1}, \quad \eta = \frac{2y}{z\bar{z} + 1}, \quad \zeta = \frac{z\bar{z} - 1}{z\bar{z} + 1}. \quad (9)$$

Interchanging primed and unprimed values in (4) and using the equation

$$r'^2 = \xi^2 + \eta^2 + (\zeta - 1)^2 = \xi^2 + \eta^2 + \zeta^2 - 2\zeta + 1 = 2(1 - \zeta),$$

we have

$$z = \frac{\xi}{1 - \zeta}, \quad y = \frac{\eta}{1 - \zeta}, \quad z = \frac{\xi + i\eta}{1 - \zeta}. \quad (10)$$

53. Rotations of the Sphere.—If Σ_0 be transformed into itself in a directly conformal manner, the corresponding points

of the z -plane undergo a one-to-one and directly conformal transformation. That is, the z -plane is subjected to a linear transformation. We shall be particularly interested in the rigid motions of space which carry Σ_0 into itself. Such a motion is, obviously, one-to-one and preserves angles.

The most general rigid motion which carries a sphere into itself is a rotation about an axis through the center. Thus, in a rigid motion carrying Σ_0 into itself, there is at least one fixed point P_1 on Σ_0 , since the corresponding linear transformation of the z -plane has a fixed point. The center O of Σ_0 is also fixed in rigid motion. Hence, the line OP_1 is a fixed axis; and the motion is a rotation.

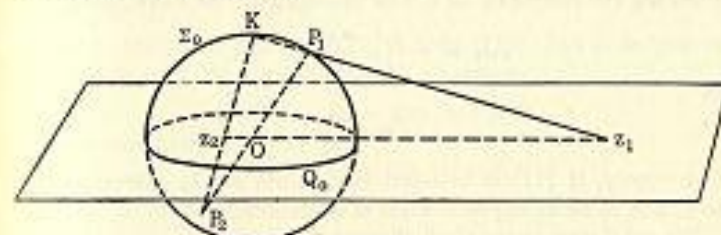


FIG. 35.

What is the character of the resulting linear transformation in the plane? Let P_1 and P_2 be the ends of the axis of rotation (Fig. 35). A great circle through P_1 and P_2 is rotated into another such circle making an angle θ (the angle through which the sphere is rotated) with the original circle. Any circle lying in a plane perpendicular to P_1P_2 is carried into itself. This latter circle is orthogonal to the circles through P_1 and P_2 . We now make the stereographic projection. P_1 and P_2 project into two points z_1, z_2 which remain fixed as the sphere is rotated. The circles through P_1 and P_2 project into circles through z_1 and z_2 . Then, each circle through z_1 and z_2 is carried into another such circle making an angle θ with its former position. The fixed circles on the sphere project into fixed circles on the plane orthogonal to the circles through z_1 and z_2 . This arrangement of the fixed circles shows that the transformation is (Fig. 7) an *elliptic transformation*.

Conversely, any elliptic transformation of the z -plane whose fixed points are the projections of the ends of a diameter corresponds to a rotation of Σ_0 . For, we can find a rotation of Σ_0

whose corresponding transformation of the z -plane has the same fixed points as the given transformation and which turns a circle through the fixed points through the same angle. The given transformation and the transformation corresponding to the rotation have the same fixed points and the same multiplier, so they are identical.

We now find the relation between two points z_1, z_2 in order that they be projections of two points P_1, P_2 which are at opposite ends of a diameter of Σ_0 . If (ξ_1, η_1, ζ_1) are the coordinates of P_1 , then $(-\xi_1, -\eta_1, -\zeta_1)$ are the coordinates of P_2 . So, from (10),

$$z_1 = \frac{\xi_1 + i\eta_1}{1 - \zeta_1}, \quad z_2 = -\frac{\xi_1 + i\eta_1}{1 + \zeta_1}.$$

Forming the conjugate of z_1 and multiplying, we have, using (8),

$$\bar{z}_1 z_2 = -\frac{\xi_1^2 + \eta_1^2}{1 - \zeta_1^2} = -1,$$

or

$$z_2 = -\frac{1}{\bar{z}_1}. \quad (11)$$

Conversely, if (11) is satisfied the points on Σ_0 corresponding to z_1 and z_2 lie at opposite ends of a diameter.

We shall now prove the following theorem:

THEOREM 1.—*The necessary and sufficient condition that a transformation of the z -plane correspond to a rotation of Σ_0 is that it be of the form*

$$z' = \frac{az - \bar{c}}{cz + \bar{a}}, \quad a\bar{a} + c\bar{c} = 1. \quad (12)$$

Let $z' = (az + b)/(cz + d)$, $ad - bc = 1$, correspond to a rotation of Σ_0 . The rotation has the property that points at the ends of a diameter remain at the ends of a diameter after the rotation. So, if z_1 and z_2 satisfy (11) so also do the transformed points $(az_1 + b)/(cz_1 + d)$ and $(az_2 + b)/(cz_2 + d)$. We have, then,

$$\frac{az_2 + b}{cz_2 + d} = -\frac{\bar{c}z_1 + \bar{b}}{\bar{a}z_1 + \bar{b}} = -\frac{\bar{d}z_1 - \bar{c}}{-\bar{b}z_1 + \bar{a}},$$

the last expression resulting from setting $\bar{z}_1 = -1/z_2$ in the preceding. From this identity in z_2 , we have

$$\frac{a}{d} = -\frac{b}{c} = -\frac{c}{b} = \frac{d}{a} = \lambda,$$

or

$$a = \lambda d, \quad b = -\lambda c, \quad c = -\lambda b, \quad d = \lambda a.$$

From these equations, we have

$$ad - bc = \lambda(\bar{a}\bar{d} - \bar{b}\bar{c}); \text{ or } 1 = \lambda.$$

Hence, $b = -\bar{c}$ and $d = \bar{a}$, and the transformation has the form (12).

We now assume that the transformation has the form (12) and prove that it corresponds to a rotation of the sphere. In the first place, (12) is elliptic, unless it is the identical transformation. If $c \neq 0$, we have $|a| < 1$, so that $|a + \bar{a}| < 2$. Since $a + \bar{a}$ is real, this is the condition that the transformation be elliptic (Theorem 15, Sec. 10). If $c = 0$, we have $|a| = 1$; then, $|a + \bar{a}| < 2$, unless $a = \pm 1$. In this latter case, we have the identity $z' = z$.

Again, the fixed points of (12) satisfy (11). Let z_1 be one of the fixed points; then z_1 satisfies the equation

$$cz_1^2 + (\bar{a} - a)z_1 + \bar{c} = 0.$$

Taking conjugates

$$\bar{c}\bar{z}_1^2 + (a - \bar{a})\bar{z}_1 + c = 0;$$

whence, dividing by $\bar{z}_1^2$,

$$\frac{c}{\bar{z}_1^2} + (a - \bar{a})\left(-\frac{1}{\bar{z}_1}\right) + \bar{c} = 0.$$

That is, $-1/\bar{z}_1$ is also a root of the equation determining the fixed points. The fixed points thus satisfy (11), and, hence, are the projections of the ends of a diameter. It follows that the transformation corresponds to a rotation of the sphere.

54. Groups of the Regular Solids.—We propose, now, to construct finite groups of linear transformations by forming finite groups of rotations of the sphere Σ_0 and projecting stereographically on the z -plane. We proceed as follows: Let one of the regular solids,—a cube, for instance, or a regular tetrahedron—be placed with its center at the center of Σ_0 . There exist certain rigid motions of space which carry the regular solid into itself. These motions, which interchange the faces in various ways, leave the center of the figure fixed. Hence, each such motion is a rotation about an axis through the center. By these motions, the sphere Σ_0 undergoes certain rotations about diametral axes; and, by stereographic projection, the z -plane is subjected to corresponding linear transformations.

It is clear that the set of all rigid motions which carry a body into itself constitutes a group of rigid motions. For, the succes-

sion of two such motions, as well as the inverse of any, is a rigid motion which also leaves the body invariant, and, hence, belongs to the set. The set of linear transformations of the z -plane corresponding to the group of rotations carrying the regular solid into itself constitutes a group of linear transformations isomorphic with the group of rotations.

It is evident that the number of rotations which carry a regular solid into itself is finite. There is but a finite number of ways, for example, in which a given face can be made to coincide with itself or with some other face of the solid. Hence, in the z -plane we have groups containing a finite number of linear transformations.¹

55. A Study of the Cube.

We begin with the most familiar of the regular solids—the cube. Let a cube be placed with its center at the origin (Fig. 36) and with its edges parallel to the coordinate axes. We may suppose that the cube is inscribed in Σ_6 . We study the axes about

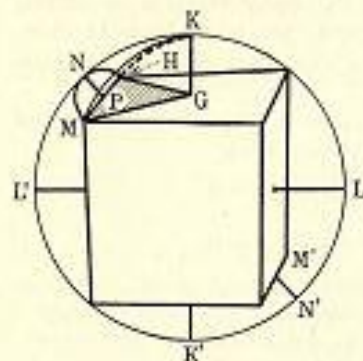


FIG. 36.

which the cube can be rotated into itself. These are of three kinds.

There are, first, the axes joining midpoints of opposite faces as KK' in the figure. There are three such axes. Rotations about each of these axes through angles of 90, 180, and 270 degrees carry the cube into itself. There are, thus, nine rotations about the three axes which carry the cube into itself.

Second, there are the axes joining opposite vertices, as MM' in the figure. There are four of these axes. Rotations about each axis through 120 and 240 degrees carry the cube into itself. There are, thus, eight rotations arising from the four axes.

Finally, there are the axes joining the midpoints of opposite edges, as NN' in the figure. There are six of these axes, a rotation about each of them through 180 degrees carries the cube into itself.

¹ All these groups are treated in KLEIN's "Vorlesungen über das Ikosaeder."

From the three types of axes we have $9 + 8 + 6 = 23$ rotations. To these we add the identical transformation; that is, the case of no rotation. The group of rotations which carry the cube into itself consists of 24 rotations.

The corresponding group in the z -plane consists of 24 linear transformations. The setting up of these transformations is a mere matter of algebra. We have, first, to find the two points where an axis of rotation meets Σ_6 . We next project the two points on the z -plane (Equation (10)). Let z_1, z_2 be the projected points. We then set up the transformation

$$\frac{z' - z_1}{z' - z_2} = e^{i\theta} \frac{z - z_1}{z - z_2}, \quad (13)$$

where θ is the angle through which the cube is rotated.

We consider a few examples from Fig. 36. The points $L(1, 0, 0)$ and $L'(-1, 0, 0)$ project into $z_1 = 1$ and $z_2 = -1$. The angle θ is $\pi/2, \pi$, or $3\pi/2$; so $e^{i\theta} = i, -1, -i$. The three transformations corresponding to the rotations about LL' are

$$\frac{z' - 1}{z' + 1} = k \frac{z - 1}{z + 1}, \quad k = i, -1, -i.$$

The points $M(-1/\sqrt{3}, -1/\sqrt{3}, 1/\sqrt{3})$ and $M'(1/\sqrt{3}, 1/\sqrt{3}, -1/\sqrt{3})$ project into, from (10), $z_1 = -\frac{1+i}{\sqrt{3}-1}, z_2 = \frac{1+i}{\sqrt{3}+1}$. Here $\theta = 2\pi/3$ or $4\pi/3$; $e^{i\theta} = \frac{1}{2}(-1 + i\sqrt{3})$ or $\frac{1}{2}(-1 - i\sqrt{3})$. On substituting in (13) we have the desired transformations.

The points $N(-1/\sqrt{2}, 0, 1/\sqrt{2})$ and $N'(1/\sqrt{2}, 0, -1/\sqrt{2})$ project into $z_1 = -(\sqrt{2} + 1)$ and $z_2 = \sqrt{2} - 1$. Here $\theta = \pi$ and $e^{i\theta} = -1$. On substituting into (13) and simplifying we have

$$z' = S(z) = -\frac{z-1}{z+1}. \quad (14)$$

The points K and K' project into ∞ and 0, respectively. The corresponding transformations in the plane are the rotations $z' = kz$, where $k = i, -1, -i$. The three rotations are powers of the transformation

$$z' = T(z) = iz. \quad (15)$$

It is evident that, if an axis passes through the midpoints of opposite faces, the corresponding three linear transformations are powers of a single transformation of period four. If the axis joins opposite vertices, the corresponding two transformations are powers of a transformation of period three. If the axis joins the midpoints of opposite edges the corresponding linear transformation is of period two. The fixed points of the

elliptic transformations which form the group are thus divided into three sets. All the fixed points of any one of the sets are congruent. Thus, K can be carried into L , K' , L' , or any point at the end of an axis of the same kind, by means of a rigid motion which carries the cube into itself. Hence, the projections of all these points are congruent. But K cannot be so carried into M or N . Similarly, the points corresponding to the ends of the axes of the kind to which MM' belongs are congruent. The same is true of the points corresponding to N , N' and similar points.

We can form a fundamental region for the group by constructing the isometric circles. We shall, however, proceed in a different manner. Consider the triangle MHG , in the figure, formed by joining the vertices at the ends of an edge to the midpoint of an adjacent face. By a suitable one of the rotations about KK' , the triangle can be carried into another such triangle abutting along the side MG , or one abutting along HG . By the rotation about NN' , we get a triangle abutting along MH . If we project the triangle MHG on the surface of the sphere, the origin being the center of projection, we have the spherical triangle formed by the arcs of great circles KM , KH , MN . This spherical triangle is a fundamental region on the sphere for the group of rotations. Its stereographic projection in the z -plane is a fundamental region for the group of linear transformations.

Let z_M , z_N , $z_{K'}$, $z_{L'}$ ($=\infty$) be the projections of M , N , H , K . Then the fundamental region has two pairs of congruent sides. The transformation T (Equation (15)) carries $z_K z_N$ into $z_K z_M$; the transformation S (Equation (14)) carries $z_N z_M$ into $z_N z_{K'}$. There are three cycles: z_K constitutes a cycle of angle $\pi/2$; z_N is a cycle of angle π ; and z_M and $z_{K'}$ constitute a cycle of angle $2\pi/3$. These cycles give the relations $T^4 = 1$ and $S^2 = 1$, which we know already, and $(ST)^2 = 1$, which is easily verified.

We can construct 23 other triangles on the cube by joining the ends of an edge to the midpoint of an adjacent face. We can carry MHG into any one of these by a suitable rotation. If these 23 triangles be projected on the sphere and then projected stereographically on the z -plane, we have the 23 regions congruent to the fundamental region constructed above.

The transformations S and T connecting congruent sides of the fundamental region are generating transformations for the group. For, by the combinations of S and T , we can construct regions adjacent to the fundamental region, regions adjacent to the new

regions, and so on as long as there are any free sides. We thus cover the whole plane; and all the transforms of the fundamental region are accounted for.

The preceding group of 24 linear transformations is known as the "octahedral group." This name is due to the fact that it is the group arising from the regular octahedron. Let the six intercepts of the sphere on the coordinate axes— K , L , K' , etc.—be joined by lines to form a regular octahedron (Fig. 37). The octahedron admits the same rotations about KK' , LL' , etc., as the cube. It is easily seen that an axis, as MM' , joining opposite vertices of the cube joins the midpoints of opposite faces of the octahedron; that an axis, as NN' , joining the midpoints of opposite edges of the cube also joins the midpoints of opposite edges of the octahedron; and that the octahedron and the cube admit the same rotations about these axes.

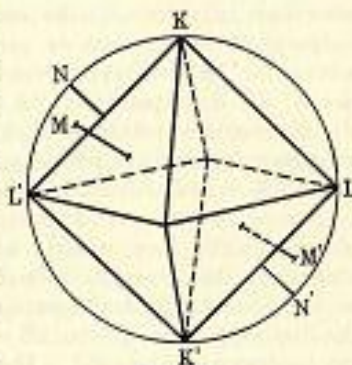


FIG. 37.

56. The General Regular Solid.—Let the number of faces of the solid be F , the number of its vertices be V , and the number of its edges be E . Let μ be the number of edges bounding each face, and let ν faces meet at the vertex.

The number of rigid motions carrying the regular solid into itself is easily found. The solid can be brought into coincidence with itself so that a given edge $a_1 b_1$ is made to coincide with any edge $a_i b_i$ or with $b_i a_i$. Then the number of rotations carrying the regular solid into itself is equal to $2E$.

Let the regular solid be inscribed in Σ_3 , and we project stereographically on the plane as in the preceding section. The various fixed points of the linear transformations in the plane are the projections of the ends of the diameters about which the rotations take place. A diameter is an axis about which the solid can be rotated into itself if, and only if, it passes through the midpoint of a face, through a vertex, or through the midpoint of an edge. The number of fixed points of the group is then

$F + V + E$. Using Euler's formula, which applies to any simply connected solid,

$$V + P = E + 2, \quad (16)$$

we have that the number of fixed points in the group is $2E + 2$.

The rotations about an axis through the midpoint of a face are all through multiples of an angle $2\pi/\mu$; that is, the rotations are powers of a rotation of period μ . The rotations about an axis through a vertex are powers of a rotation of period ν . The rotation about an axis through the midpoint of a side is of period 2. These statements are true for the corresponding transformations in the plane.

A midpoint of a face can be carried into the midpoint of any other face; a vertex can be carried into any other vertex; a midpoint of an edge can be carried into the midpoint of any other edge by a rotation which carries the solid into itself. Hence, the fixed points of the plane are separated into three sets of congruent points: F fixed points are congruent; V others are congruent; and the remaining E are congruent.

A fundamental region for the group can be got, as in Fig. 36, by joining the vertices at the ends of an edge to the midpoint of an adjacent face to form a triangle; then projecting this triangle on the sphere; and, thence, projecting stereographically on the plane. The transforms of this fundamental region are the like projections of the $2E - 1$ other such triangles which can be drawn on the solid. The congruent sides of this fundamental region are connected by a transformation T of period μ and a transformation S of period 2. S and T are generating transformations for the group. They are connected by the relation $(ST)^2 = 1$.

In the following table, the various regular solids are listed with the number of their faces, vertices, etc. In the last column $N (-2E)$ is the number of transformations of the group:

TABLE I

	F	V	E	μ	ν	N
Tetrahedron.....	4	4	6	3	3	12
Cube.....	6	8	12	4	3	24
Octahedron.....	8	6	12	3	4	24
Dodecahedron.....	12	20	30	5	3	60
Icosahedron.....	20	12	30	3	5	60
Dihedron.....	2	∞	∞	∞	2	2∞

The dihedron is an ideal solid of zero volume. Let a regular polygon of n sides be inscribed in Q_0 , the equator of Σ_0 . We shall look upon this figure as a regular solid with two coincident faces. The line KK' is an axis about which the figure can be rotated into itself through multiples of the angles $2\pi/n$. Also, any diameter through a vertex or through the midpoint of a side of the polygon is an axis about which the figure can be rotated through the angle π into itself. The group of rotations of this figure also carries into itself the double pyramid formed by joining K and K' to the vertices of the polygon inscribed in Q_0 . This latter figure, however, is not a regular solid.

We note the same sort of duality between the dodecahedron and the icosahedron that exists between the cube and the octahedron. If we interchange the numbers F and V and the numbers μ and ν in the tabulated values for the dodecahedron, we have the entries for the icosahedron. It is not difficult to verify the fact that the icosahedron can be so placed as to have precisely the same group of rotations as the dodecahedron. The axes joining the midpoints of the opposite faces of one solid are made the axes joining the opposite vertices of the other.

Two further groups remain to be mentioned. The *four group* corresponds to the four rotations of the sphere Σ_0 which carry the real axis into itself and the imaginary axis into itself. It may be regarded as a limiting case of the dihedral group in which $n = 2$.

There is, finally, the simplest of all the finite groups—the elliptic cyclic group. This group has two non-congruent fixed points. We can give it a geometrical origin, if we like, by considering it as arising from the group of rotations which carries into itself a regular pyramid formed by joining the vertices of a regular polygon inscribed in Q_0 to the point K .

We have found five types of finite groups of linear transformations: the *elliptic cyclic group*, the *dihedral group* (including the *four group*), the *tetrahedral group*, the *octahedral group*, and the *icosahedral group*. Each of these groups can be transformed after the manner of Sec. 15, to get other finite groups. The possibility of further finite groups will be investigated in the next section. We shall find that there are no others.

57. Determination of All the Finite Groups.—We first prove the following theorem:

THEOREM 2.—If a group contains two non-parabolic transformations which have one fixed point in common and the other fixed

points different, then the group contains parabolic transformations with the common point as fixed point.

Let the group be so transformed that ∞ is the common fixed point and 0 and 1 are the remaining fixed points. The two transformations are then of the form

$$S_1 = K_1 z, \quad S_2 = K_2(z-1) + 1, \quad K_1 \neq 1, K_2 \neq 1.$$

Combining these in the product given below, we find readily

$$U = S_1^{-1} S_2 S_1 = z + \frac{(K_1 - 1)(K_2 - 1)}{K_1}.$$

This is a translation; that is, a parabolic transformation with ∞ as fixed point. Any power of this transformation U^n , $n \neq 0$, is parabolic with ∞ as fixed point.

A finite group has only elliptic transformations. It follows from the preceding theorem that all transformations of the group which have one fixed point in common have the second fixed point in common, also. If ν transformations (including the identical transformation) have a common fixed point, e , we shall call e a fixed point of order ν . These transformations form a cyclic subgroup; and the transformations are powers of a transformation with multiplier $e^{2\pi i/\nu}$. Congruent fixed points are of the same order. Thus, if $e' = T(e)$; and S_i , $i = 1, 2, \dots$ are the transformations with e as fixed point, then the ν transformations $TS_i T^{-1}$, and no others, have e' as fixed point.

Given a finite group of $N(>1)$ linear transformations, so transformed, if necessary, that ∞ is not a fixed point for any transformation. Let R be the fundamental region exterior to all the isometric circles. We shall investigate the cycles of the region.

Let z_0 be a vertex belonging to a cycle of angle $2\pi/k$, $k > 1$. Then we found in Sec. 26 that z_0 is a fixed point of an elliptic transformation with multiplier $e^{2\pi i/k}$. There is no transformation with z_0 as fixed point and with multiplier $e^{i\theta}$ where $0 < \theta < 2\pi/k$; so z_0 is a fixed point of order k .

Consider now a fixed point e of order ν lying outside R . There is a point e' congruent to e lying within or on the boundary of R . But e' is the fixed point of an elliptic transformation of order ν , and has isometric circles passing through it. Hence, e' is a vertex of R . That is, each fixed point of order ν is congruent to a vertex of R belonging to a cycle of angle $2\pi/\nu$.

Let a, b be finite inner points of R . We form the function

$$f(z) = \frac{(z-a)(z_1-a) \cdots (z_{\nu-1}-a)}{(z-b)(z_1-b) \cdots (z_{\nu-1}-b)} \quad (17)$$

where $z_1, z_2, \dots, z_{\nu-1}$ are congruent to z . This function is automorphic with respect to the given group, since if z be transformed into z_1 the variables $z, \dots, z_{\nu-1}$ are permuted and $f(z)$ is unaltered. It has zeros of the first order at a and the points congruent thereto and poles of the first order at b and congruent points. It has a single pole in R ; hence, it takes on every value once in R (Sec. 42, Theorem 11).

The proposition that an automorphic function takes on every value the same number of times in the fundamental region requires that we count the values in a particular way at the vertices. Let $f(z_0) = A$; and let $f(z)$ take on the value A r times. We count the value A as taken on r/k times in the vertices of the cycle (Sec. 42 (2)). In the present case, $r/k = 1$; whence, $f(z)$ takes on the value A exactly k times at z_0 . The derivative, $f'(z)$, has $k-1$ zeros at z_0 .

Now let us look upon $f(z)$ as a rational function in the whole plane. It has N poles; so it takes on each value N times. It takes on the value A only at the vertex z_0 and the points congruent thereto; and at each such point it takes on the value k times. It follows that the point z_0 and the points congruent to it form a set of N/k congruent points. Since any fixed point is congruent to a vertex of R , we can state this result as follows: *A fixed point of order ν belongs to a set of N/ν congruent fixed points. The order of a fixed point is a submultiple of N .*

Consider $f'(z)$. Since $f(z)$ has N poles of the first order, its derivative has N poles of the second order. Hence, $f'(z)$ has $2N$ zeros. The roots of $f'(z)$ can all be located. At a finite point which is different from a fixed point, $f'(z) \neq 0$. Otherwise, $f(z)$ would take on its value twice at the point; this is contrary to the fact that $f(z)$ takes on no value more than once in R or in any of the regions congruent to R which cover the plane. At infinity we have developments of the form

$$f(z) = c_0 + \frac{c_1}{z} + \cdots; \quad f'(z) = -\frac{c_1}{z^2} + \cdots$$

Here, $c_1 \neq 0$, since $f(z)$ takes on the value c_1 only once at infinity. Hence, $f'(z)$ has a zero of the second order at infinity. The

remaining zeros of $f'(z)$, $2N - 2$ in number, are at the fixed points.

Let the fixed points of the group fall into s sets of congruent points. Let the orders of the points of the sets be $\nu_1, \nu_2, \dots, \nu_s$. Then the sets consist of $N/\nu_1, N/\nu_2, \dots, N/\nu_s$ points, respectively. Summing the zeros of $f'(z)$ and equating to $2N - 2$, we have

$$\sum_{i=1}^s \frac{N}{\nu_i} (\nu_i - 1) = 2N - 2,$$

or

$$\sum_{i=1}^s \left(1 - \frac{1}{\nu_i}\right) = 2 - \frac{2}{N}. \quad (18)$$

The integers N and ν_i for any finite group must satisfy (18). Here $N \geq 2$ and $\nu_i \geq 2$; also N/ν_i is an integer. We have $s > 1$; for if $s = 1$, the first member of (18) is less than 1 and the second is greater than or equal to 1. Also, $s < 4$; for $1 - \frac{1}{\nu_1} \geq \frac{1}{2}$ and the first member equals or exceeds $s/2$, which is greater than the second member if $s \geq 4$. There are then two cases to consider $s = 2$ and $s = 3$.

If $s = 2$, (18) becomes

$$1 - \frac{1}{\nu_1} + 1 - \frac{1}{\nu_2} = 2 - \frac{2}{N}, \text{ whence, } \frac{N}{\nu_1} + \frac{N}{\nu_2} = 2.$$

We have $N/\nu_1 = N/\nu_2 = 1$. Hence, N may be any integer and $\nu_1 = \nu_2 = N$. The two sets of congruent fixed points contain one point each; and each fixed point is of order N . The groups satisfying these conditions, obviously, consist of the cyclic groups of N transformations.

If $s = 3$, (18) reduces to

$$\frac{1}{\nu_1} + \frac{1}{\nu_2} + \frac{1}{\nu_3} = 1 + \frac{2}{N}. \quad (19)$$

Let the subscripts be so chosen that $\nu_1 \leq \nu_2 \leq \nu_3$. Then $\nu_1 = 2$, for, otherwise, the first member does not exceed 1. Then,

$$\frac{1}{\nu_2} + \frac{1}{\nu_3} = \frac{1}{2} + \frac{2}{N}. \quad (20)$$

Then, $\nu_2 = 2$ or $\nu_2 = 3$; otherwise, the first member does not exceed $\frac{1}{2}$.

If $\nu_2 = 2$, we have $\nu_3 = N/2$. If N is any even number, $N = 2n$, the equation can be satisfied.

If $\nu_2 = 3$, we have, from (20),

$$\frac{1}{\nu_3} = \frac{1}{6} + \frac{2}{N} \quad (21)$$

whence, $\nu_3 < 6$. We find that each of the possible values $\nu_3 = 3, 4, 5$ gives an integral value of N ; namely $N = 12, 24, 60$, respectively.

We have found then that, except for the cyclic group ($s = 2$), all finite groups have three sets of congruent fixed points. The orders of the points, the numbers in each set, and the total number of transformations are only such as appear in the following table:

TABLE II

ν_1	ν_2	ν_3	N/ν_1	N/ν_2	N/ν_3	N
2	2	n	n	n	2	$2n$
2	3	3	6	4	4	12
2	3	4	12	8	6	24
2	3	5	30	20	12	60

That there actually exist finite groups corresponding to the possibilities set forth in Table II is evident on an inspection of Table I. In that table, there are three sets of congruent fixed points containing P , V , and E points. The corresponding orders are μ , ν , and 2, respectively. The first possibility in Table II is realized in the dihedral group (in the four group, if $n = 2$), the second in the tetrahedral group, the third in the octahedral group, and the last in the icosahedral group.

THEOREM 3.—*There are no finite groups of linear transformations other than the elliptic-cyclic groups, the groups of the regular solids (including the four group), and the transforms of these latter groups by means of linear transformations.*

We have found that when $s = 2$ the only groups arising are the cyclic groups. We shall now show that any group with three sets of congruent fixed points is the transform of a group arising from one of the regular solids. We shall prove (which amounts to the same thing) that if two groups have the same values of ν_1, ν_2, ν_3 one is the transform of the other.

Given two finite groups with the same values of ν_1, ν_2, ν_3 . Let each be so transformed, if necessary, that ∞ is not a fixed point. If the transformed groups are the transforms of one another, the same is true of the original groups. Let the group S consist of the transformations $S_1, \dots, S_N$ and have the fixed points $a_1, a_2, \dots, a_{N/\nu_1}$ of order ν_1 , the fixed points $b_1, b_2, \dots, b_{N/\nu_2}$ of order ν_2 , and the fixed points $c_1, c_2, \dots, c_{N/\nu_3}$ of order ν_3 . Let the group S' have the transformations $S'_1, \dots, S'_{N'}$, the fixed points $a'_1, \dots, a'_{N'/\nu_1}$ of order ν_1 , the fixed points $b'_1, \dots, b'_{N'/\nu_2}$ of order ν_2 , and the fixed points $c'_1, \dots, c'_{N'/\nu_3}$ of order ν_3 .

We now use formula (17) to set up an automorphic function $f(z)$ for the group S and an automorphic function $f_1(z)$ for the group S' where the constants a and b in the formula are different from the fixed points. Let $f(a_i) = A, f(b_i) = B, f(c_i) = C; f_1(a'_i) = A', f_1(b'_i) = B', f_1(c'_i) = C'$. Let $t = (at_1 + b)/(ct_1 + d)$ be the transformation carrying the distinct points $t_1 = A', B', C'$ into the distinct points A, B, C , respectively, and form the function

$$F(z) = \frac{af_1(z) + b}{cf_1(z) + d}.$$

Then, $F(a'_i) = A, F(b'_i) = B, F(c'_i) = C$. The functions $F(z)$ and $f(z)$ have N poles each, and, hence, take on any value N times.

We shall represent the group S in the z -plane and S' in a second or z' -plane. We shall set up a correspondence between the points of the two planes by means of the equation¹

$$F(z') = f(z). \quad (22)$$

To each value of z there correspond N values of z' ; for a given z determines one value of $f(z)$ and this value is taken on by $F(z')$ at N points z' . Likewise, to each z' correspond N values of z .

If $z \neq a_i, b_i, c_i$ then $f(z)$ is distinct from A, B, C and the corresponding values of z' are N distinct points in the z' -plane.

If $z = a_i$, then $f(z) = A$ and the corresponding values of z' fall into N/ν_1 sets of ν_1 equal values each; namely, $z' = a'_1, a'_2, \dots$. Consider the arrangement of the ν_1 values of z' in the neighborhood of a'_1 when z is in the neighborhood of a_i . We have the following developments in the neighborhoods of

¹ This correspondence was suggested by Prof. P. Koebe.

the points in question, since each function takes on its value ν_1 times:

$$\begin{aligned} f(z) &= A + k(z - a_i)^{\nu_1} + \dots, & k \neq 0. \\ F(z') &= A + k'(z' - a'_1)^{\nu_1} + \dots, & k' \neq 0. \end{aligned}$$

Then, (22) becomes

$$(z' - a'_1)^{\nu_1} (k' + \dots) = (z - a_i)^{\nu_1} (k + \dots).$$

Extracting the ν_1 -th root,

$$(z' - a'_1) (k'^{1/\nu_1} + \dots) = \epsilon_k (z - a_i) (k^{1/\nu_1} + \dots),$$

$$\epsilon_k = 1, 2, \dots, \nu_1,$$

where $\epsilon_1, \dots, \epsilon_{\nu_1}$ are the ν_1 -th roots of unity. From these we have the ν_1 developments in the neighborhood of a_i ,

$$z' = a'_1 + \epsilon_k \left(\frac{k}{k'} \right)^{1/\nu_1} (z - a_i) + \dots$$

Hence, although ν_1 values of z' become equal at a_i , these values belong to distinct branches. Similar reasoning applies to b_i and c_i .

The relation (22), then, gives in the neighborhood of any point of the z -plane N distinct function elements. These elements combine to form N single-valued functions of z . Otherwise, since there are no branch points, the N sheets bearing the values of z' are not connected. Consider one of the functions, $z' = T(z)$. To each value of z there corresponds one and only one value of z' . Interchanging the rôles of z and z' , to each z' there corresponds one, and only one, z . Hence, $z' = T(z)$ is a linear transformation (Sec. 1, Theorem 3, Corollary 1).

The transformation T carries the fixed points of the group S into the fixed points of the group S' ; since when $z = a_i, b_i, c_i$ then $T(z) = a'_1, b'_1, c'_1$. Then the group $U_i = TS_iT^{-1}$ is a group with the same fixed points and of the same orders as the group S' . Is the group U the same as the group S' ?

Let U_i and S'_j be transformations with the fixed point a'_1 and the same multiplier. We shall show that the second fixed points are identical. Suppose they are different. We combine U_i and S'_j in all possible ways to generate a group Γ . A point $a'_1 (\neq a'_i)$ is carried by each of the transformations into another of the points a'_i ; that is, a'_1 has a finite number of distinct transforms by the group Γ . But Γ contains parabolic transformations with a'_1 as fixed point (Theorem 2); and by

repetitions of such a parabolic transformation a_i' has an infinite number of distinct transforms. This contradiction proves that U_i and S_i' have the same fixed points; whence they are identical. We thus identify each transformation of U with a transformation of S' . The group S' is the transform of S by a linear transformation, which was to be proved.

The Polyhedral Functions.—The automorphic functions belonging to the finite groups, which, owing to their connection with the regular solids, are called "polyhedral functions," are readily set up. The function $f(z)$ in (17) takes on each value once in the fundamental region. Hence (Sec. 43, Theorem 14), the most general simple automorphic function belonging to the group is a rational function of $f(z)$.

58. The Extended Groups.—Intimately connected with the rotations which carry a regular solid into itself are the reflections

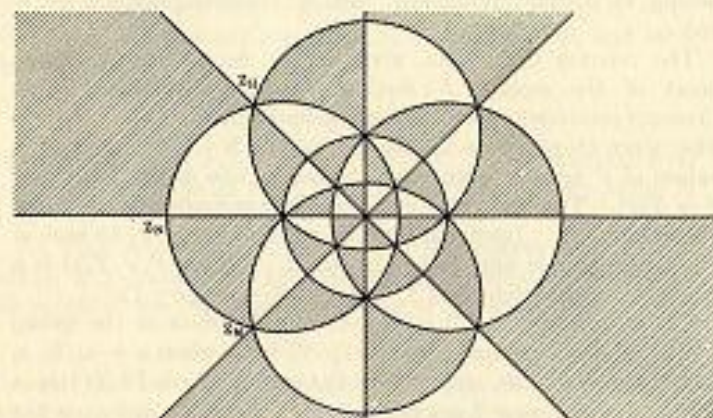


FIG. 38

in planes of symmetry which, likewise, carry the solid into itself. In the cube, Fig. 36, there are nine planes of symmetry: three planes through the midpoints of four faces (the coordinate planes), and six planes containing opposite edges. These planes intersect the sphere Σ_s in nine great circles. These circles are projected stereographically on the z -plane in Fig. 38. The real and imaginary axes and the unit circle Q_1 arise from the first

three planes of symmetry mentioned above; the remaining two straight lines and four circles arise from the latter six planes.

What is the transformation of the z -plane corresponding to a reflection in a diametral plane of the points of Σ_s ? On making the inversion which projects Σ_s stereographically on the z -plane, a diametral plane Π through a great circle C is carried into a sphere Σ through C' , the stereographic projection of C . Since Π is orthogonal to Σ_s , Σ is orthogonal to the z -plane and has C' as equator. Two points on Σ_s inverse with respect to Π are carried into two points of the z -plane inverse with respect to Σ and, hence, inverse with respect to C' . Hence, if the points of Σ_s be transformed by a reflection in Π the corresponding points of the z -plane undergo an inversion in C' . Reflection in the nine planes of symmetry of the cube, then, correspond to inversions in the nine circles of Fig. 38.

The planes of symmetry divide the surface of the cube into 48 equal triangles. One such is the triangle GHP formed by bisecting the triangle MHG by a line from G to the midpoint of MH . This triangle projects on the triangle KNH on the sphere and the latter projects stereographically on the shaded triangle in the upper left hand corner of Fig. 38. The triangle GPM corresponds to the unshaded region in the lower left-hand corner of Fig. 38; and the remaining triangles on the cube correspond to the remaining triangles of Fig. 38.

Now, let us form a group by combining the reflections in planes of symmetry in all possible ways. Each such reflection carries GHP into one of the 48 triangles on the surface of the cube. Further, by suitable sequences of such reflections, the whole surface of the cube can be covered; that is, GHP can be carried into any other triangle. It is easily seen that two such sequences which carry GHP into the same triangle transform all points in the same way and so are identical. The group thus contains exactly 48 transformations. In the z -plane the corresponding transformations carry a given triangle of Fig. 38 into each of the triangles of the figure.

This group in the z -plane is known as the "extended group." It is found by combining in all ways the inversions in the nine circles in Fig. 38. Each triangle is a fundamental region for the group. The transformations are of two kinds. An even number of inversions is a linear transformation; an odd number is an inversely conformal transformation of the plane into itself,

and has the form $z' = (az + b)/(cz + d)$ (Sec. 5, Equation (23)). There are 24 transformations of each kind. The former correspond to the rotations of the sphere which carry the cube into itself, since an even number of reflections in diametral planes is equivalent to a rigid motion. Among the inversely conformal transformations are the nine inversions from which we formed the group.

If we shade the alternate triangles in Fig. 38, we note that a transformation which carries a shaded triangle into a shaded triangle or an unshaded triangle into an unshaded triangle is a linear transformation; a transformation which carries a shaded triangle into an unshaded triangle or vice versa is a transformation which reverses the sign of the angle. The former 24 transformations constitute the octahedral group. A shaded and an unshaded region together constitute a fundamental region for the group.

It is easy to see that the extended group can be generated by three inversions; namely, in the sides of any one of the triangles of the figure. By repeating these three inversions, any triangle of the figure can be carried into any other.

The groups of the remaining regular solids can be extended in a precisely similar way. We shall not treat the several cases in detail. By repeated reflections in planes of symmetry, a given edge a_0b_0 can be carried into an edge a_ib_i in four ways—so that a_0b_0 coincides with a_ib_i or with b_0a_0 , and so that angles are preserved or angles are reversed. The extended group corresponding to the sequence of reflections then consists of $4E$ transformations, where E is the number of edges.

The planes of symmetry divide the surface of the regular solid into $4E$ triangles, each formed by joining the midpoint of a face to the end and the midpoint of an adjacent edge. The projection of these triangles on the sphere and thence on the z -plane gives a system of triangles with properties analogous to those of Fig. 38. From the tetrahedron we get 24 triangles; from the icosahedron 120; from the dihedron $4n$. In each case, half the transformations are linear transformations and half are transformations with reversal of the sign of the angle. The former constitute a subgroup identical with the group of the regular solid.

The extended groups will reappear in a later chapter in the study of the Riemann-Schwarz triangle functions.

II. THE GROUPS WITH ONE LIMIT POINT

59. The Simply and Doubly Periodic Groups.—The groups with one limit point are of particular interest. They comprise the groups whose automorphic functions—the simply and doubly periodic functions—have been studied more than any other automorphic functions. Further, as we shall find later, these groups form one of the three classes of groups whose automorphic functions are employed in the simplest theory of uniformization.

A group with but one limit point can contain no hyperbolic or loxodromic transformation; for each of the two fixed points of such a transformation is a limit point. Each transformation of the group is either parabolic or elliptic. Each transformation has the limit point as fixed point; otherwise, the limit point would have a transform which would be a second limit point. The group cannot consist entirely of elliptic transformations. For, the second fixed points do not all coincide; otherwise, the group would be a finite cyclic group. It follows from Theorem 2 that the group contains parabolic transformations. The group, then, consists either of parabolic transformations alone or of parabolic and elliptic transformations.

Let the group be so transformed that the limit point is carried to infinity. Then all the transformations are of the form

$$z' = Kz + b,$$

where K is the multiplier of the transformation. The parabolic transformations ($K = 1$) are translations; the elliptic transformations are rotations.

If the group contains only translations, it is either a *simply periodic group* with a period strip (Fig. 12) as fundamental region or a *doubly periodic group* with a period parallelogram (Fig. 13) as fundamental region.

We prove these facts briefly. The translations are of the form $S_i = z + \Omega_i$, Ω_i being called a "period." We have $S_1 S_2 S_1^{-1} \cdots S_n S_n^{-1} = z + m_1 \Omega_1 + m_2 \Omega_2 + \cdots + m_n \Omega_n$; that is, any linear combination of periods with integral coefficients is a period. Now, let all the periods be plotted in the complex plane. These points have no cluster point. If there were a cluster point, the transformation $S_i S_i^{-1} = z + \Omega_i - \Omega_i$, where Ω_i and Ω_i are taken sufficiently near the cluster point, would have an arbitrarily small period; and the group would be continuous.

Let ω be a nearest period to the origin; that is, one of the finite number at the minimum distance. Let L be the line joining the origin to ω and let Ω be any period lying on L . We can write $\Omega = M\omega$, where M is real. We

shall show that M is an integer. Suppose not; and let m be the nearest integer to M . Then $|M - m| < \frac{1}{2}$; and the period $\Omega' = \Omega - m\omega$ is such that $|\Omega'| = |(M - m)\omega| < \frac{1}{2}|\omega|$. This is contrary to the assumption that ω is the nearest period to the origin. The transformation $S = z + \Omega$ is then a power of $T = z + \omega$; namely, $S = T^m$. If all the periods lie on L , the group is the simply periodic group generated by T .

Suppose, next, that there are periods not lying on L . Let ω' be at the minimum distance from the origin among these. Then any period in the plane can be written $\Omega = M\omega + M'\omega'$, where M and M' are real. We now prove that M and M' are both integers. Suppose not; and let m and m' be the nearest integers to M , M' . Then $|M - m| < \frac{1}{2}$, $|M' - m'| < \frac{1}{2}$. Consider the period

$$\Omega' = \Omega - (m\omega + m'\omega') = (M - m)\omega + (M' - m')\omega' \neq 0.$$

Here we have

$$|\Omega'| < \frac{1}{2}|\omega| + \frac{1}{2}|\omega'| \leq |\omega'|.$$

This is impossible unless Ω' lies on L . But, then, $M' - m' = 0$ and $|\Omega'| < \frac{1}{2}|\omega|$, which is impossible. It follows that M and M' are integers.

A transformation of the group, $S = z + \Omega$, can then be written in the form $S = T^m T_1^{m'}$, where $T = z + \omega$, $T_1 = z + \omega'$. The group is the doubly periodic group generated by T and T_1 .

We can assume, without loss of generality, that a period of smallest absolute value is 1. Let $G = z/\omega$, where ω is a smallest period, and transform the group by G .

$$S = z + \Omega, \quad GSG^{-1} = G(\omega z + \Omega) = z + \frac{\Omega}{\omega}.$$

When $\Omega = \omega$, the period is 1. Otherwise, $|\Omega/\omega| \geq 1$, since $|\Omega| \geq |\omega|$.

The most general group with one limit point and containing only parabolic transformations is the transform of a simply or doubly periodic group.

60. Groups Allied to the Periodic Groups.¹—We now consider groups containing parabolic and elliptic transformations. We show first that the multiplier of an elliptic transformation is limited to a small number of possible values. Let $S = z + \Omega$ and $S_1 = Kz + b$ be a parabolic and an elliptic transformation contained in the group. Then the translation

$$S_1 S S_1^{-1} = S_1 \left(\frac{z - b}{K} + \Omega \right) = z + K\Omega$$

has the period $K\Omega$. Let the minimum period be 1; then, taking $\Omega = 1$, K is a period. Now, the multipliers of the transformations with a common fixed point of order ν are

$$K_\nu (= e^{2\pi i/\nu}), K_{\nu-1}, \dots, K_1 (= 1).$$

¹ For the treatment in this section and the next see KOSAR, P., *Math. Ann.*, vol. 67, pp. 164-168, 1909.

These multipliers—the ν -th roots of unity—lie on the unit circle Q_ν .

The possible values of K and the periods ± 1 must be so spaced about Q_ν that the distance between any two distinct periods is not less than 1. Otherwise, there is a period of smaller absolute value than 1, contrary to hypothesis. If $\nu \geq 7$, (Fig. 39) we have $|K_\nu - 1| < 1$, which is impossible. If $\nu = 5$, we have $|K_2 + 1| < 1$, which is impossible. Hence, ν is limited to the values 2, 3, 4, 6. Also $\nu = 4$ and $\nu = 3$, or $\nu = 6$, cannot both appear; since $|K_4 - K_3| < 1$ and $|K_4 - K_6| < 1$.

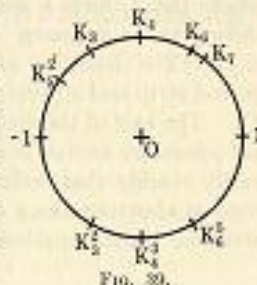


FIG. 39.

Two elliptic transformations with the same multiplier, $S_1 = Kz + b$, $S_2 = Kz + b'$, have constants b , b' which differ by a period. For,

$$S_2 S_1^{-1} = K \left(\frac{z - b}{K} \right) + b' = z + b' - b,$$

which has the period $b' - b$. Conversely, if Ω is any period, there is a transformation with the multiplier K and the constant $b + \Omega$; namely, where $S = z + \Omega$,

$$SS_1 = Kz + b + \Omega.$$

Then all transformations with the multiplier K which the group contains, and no others, are comprised in the formula

$$S_i = Kz + b + \Omega_i, \quad (23)$$

where b is the constant for one such transformation and Ω_i is any period or zero. The finite fixed point of (23) we find to be

$$\xi_i = \frac{b + \Omega_i}{1 - K}. \quad (24)$$

In locating the fixed points from (24), it suffices to find those which lie in a period strip or period parallelogram of the subgroup of translations. The remaining fixed points differ from these by periods. For, if U is any transformation of the group, US_iU^{-1} has the same multiplier as S_i and has the fixed point $U(\xi_i)$. In particular, each fixed point has a congruent fixed point lying in the period strip or period parallelogram.

We treat first the case in which $\nu = 2$ for all the fixed points; so $K = -1$. We can transform the group by a translation

carrying b to the origin; so we can take $b = 0$. The resulting transformations,

$$S_i = -z + \Omega_i, \quad T_i = z + \Omega_i,$$

where the T_i form a simply or doubly periodic group, are easily shown to be a group. The fixed points, from (24), are $\xi_i = \Omega_i/2$. The location of the fixed points (marked "2") in a period strip and a period parallelogram are shown in Figs. 40 and 41. The half of the strip or parallelogram whose congruent sides are joined by arrows is a fundamental region for the group. We verify readily that certain transformations carry the region into regions abutting along each of its sides, and that all except the identical transformation carry it outside itself.

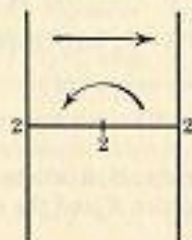


FIG. 40.

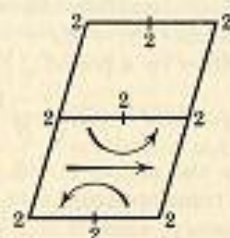


FIG. 41.

If $\nu > 2$, the subgroup of translations cannot be simply periodic. For K_ν is then a period which is not a multiple of 1. Since K_ν is at a distance 1 from the origin and no period can be at a less distance, we may, according to the previous section, take K_ν as the period of the second generating transformation. The doubly periodic subgroup is generated by $T = z + 1$ and $T_1 = z + K_\nu$. The doubly periodic group is, thus, a very special one.

Consider the case $\nu = 4$. We have $K_4 = i$, $K_4^2 = -1 = K_2$, $K_4^3 = -i$. Since $K_4^4 = 1$, we may write the transformations, both elliptic and parabolic, in the form

$$S_i = K_4^i z + \Omega_i = K_4^i z + m + m'i.$$

Both group properties are readily established for these transformations.

For $K = i$ and $K = -1$ we have from (24)

$$\xi_i = \frac{m + m'i}{1 - i} = \frac{(1 + i)(m + m'i)}{2}, \quad \xi_i = \frac{m + m'i}{2}$$

The first formula gives the points marked "4" in the period parallelogram in Fig. 42. The second gives those marked "4" and in addition those marked "2." These latter points are of order 2. The fixed points when $K = -i$ are, of course, the same as those for $K = i$. We can show easily that the square whose congruent sides are connected by arrows is a fundamental region for the group.

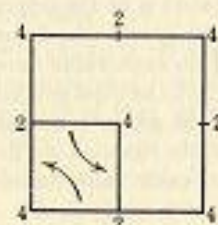


FIG. 42.

The cases $\nu = 3$ and $\nu = 6$ are pictured in Figs. 43 and 44. The group in each case is

$$S_i = K_\nu^i z + m + m'K_\nu, \quad \nu = 1, 2, \dots, \nu.$$

The fixed points are located as in the preceding figure. In the latter case there are fixed points of three orders, 2, 3, and 6. In Fig. 43, the parallelogram has been drawn from the periods

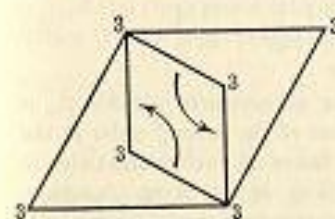


FIG. 43.

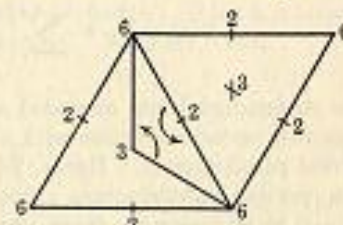


FIG. 44.

1 and $K_3 + 1 (= K_6)$; this latter period is at a distance 1 from the origin and may be used, as well as K_3 itself, as the period of the second generating transformation.

We observe that in Figs. 42, 43, and 44, the transformations connecting congruent sides of the fundamental region are elliptic. Each of these groups is generated by rotations.

The groups constructed in this section, together with their transforms by means of linear transformations, include all

the groups with a single limit point which contain elliptic transformations.

61. The Automorphic Functions.—We shall now set up all the simple automorphic functions for the groups with one limit point. We suppose the group so transformed that the limit point is at infinity, the smallest period is 1, and an elliptic point of highest order (if elliptic points exist) is at the origin.

For the simply periodic group, the function $e^{2\pi iz}$ is an automorphic function taking on each value once in the period strip. Hence (Sec. 43, Theorem 14), the most general simple automorphic function connected with the group is a rational function of $e^{2\pi iz}$.

In the group of Fig. 40, the function $\cos 2\pi z$ is a simply periodic function which takes on each value twice in the period strip (since it has two zeros, $z = \pm \frac{1}{4}$, in the strip). Now $\cos 2\pi(-z) = \cos 2\pi z$. The function thus admits the elliptic generating transformation; so it is automorphic with respect to the group of the figure. It takes on each value once only in the upper half of the strip, the other value being taken on in the lower half. According to Theorem 14, cited above, the most general simple automorphic function connected with the group is a rational function of $\cos 2\pi z$.

The group of Fig. 41 is treated similarly. The Weierstrassian function

$$\wp(z) = \frac{1}{z^2} + \sum_i \left[\frac{1}{(z - \Omega_i)^2} - \frac{1}{\Omega_i^2} \right], \quad (25)$$

the summation being extended over all non-zero periods Ω_i , is a doubly periodic function with a pole of the second order in the period parallelogram. Hence, $\wp(z)$ takes on each value twice in the period parallelogram. But $\wp(z)$ is an even function, $\wp(-z) = \wp(z)$ —as we see on changing z to $-z$ and Ω_i to $-\Omega_i$, which does not alter the set of periods. Then $\wp(z)$ takes on each value once in the fundamental region of the figure. The most general simple automorphic group is a rational function of $\wp(z)$.

For the doubly periodic function itself there is no automorphic function with a single pole in the period parallelogram in terms of which to express all automorphic functions rationally. Let $f(z)$ be a simple automorphic function belonging to the group; that is, an elliptic function. Then $f(z)$ can be written in the form

$$f(z) = \frac{1}{2}[f(z) + f(-z)] + \frac{1}{2}[f(z) - f(-z)].$$

Here, $f(-z)$ is also an elliptic function. The first term is an even function of z ; and so is automorphic with respect to the group of Fig. 41. It is a rational function of $\wp(z)$. The second term is an odd function. Now $\wp'(z)$ is also an odd elliptic function. Hence, $\frac{1}{2}[f(z) - f(-z)]/\wp'(z)$ is an even function and so is a rational function of $\wp(z)$. The most general simple automorphic function is then

$$f(z) = R_1[\wp(z)] + \wp'(z)R_2[\wp(z)],$$

where R_1 and R_2 indicate rational functions.

Consider now the groups of Figs. 42 to 44. The transformations here are of the form

$$z_i = K_\nu z + \Omega_i, \quad i = 1, 2, \dots, \nu, \quad (26)$$

where $\nu = 4, 3$, or 6 . The function $\wp(z)$ is not automorphic with respect to the groups. We have, from (25),

$$\wp(z_i) = \frac{1}{(K_\nu z + \Omega_i)^2} + \sum_i' \left[\frac{1}{(K_\nu z + \Omega_i - \Omega_i')^2} - \frac{1}{\Omega_i'^2} \right] \\ = \frac{1}{K_\nu^2} \left[\frac{1}{(z + \frac{\Omega_i}{K_\nu})^2} + \sum_i' \left[\frac{1}{(z + \frac{\Omega_i - \Omega_i'}{K_\nu})^2} - \frac{1}{(\frac{\Omega_i'}{K_\nu})^2} \right] \right].$$

Now, the division of a period Ω_i by K_ν amounts to rotating it through the angle $-2\pi\nu/\nu$ about the origin and carries it in each of the three cases into another period Ω_i' . Such a rotation carries the set of all periods into itself. We have, then,

$$\wp(z_i) = \frac{1}{K_\nu^2} \left[\frac{1}{(z + \Omega_i')^2} + \sum_i' \left[\frac{1}{(z + \Omega_i' - \Omega_i'')^2} - \frac{1}{\Omega_i''^2} \right] \right] \\ = \frac{1}{K_\nu^2} \wp(z + \Omega_i') = \frac{1}{K_\nu^2} \wp(z). \quad (27)$$

For $\nu = 4$, (Fig. 42) $K = \pm i$, ± 1 and $K_\nu^2 = \pm 1$. Then, $\wp(z)^2$ is automorphic with respect to the group. $\wp(z)^2$ has a pole of the fourth order in the period parallelogram and takes on each value four times. The period parallelogram contains four copies of the fundamental region in the figure. Hence, the function takes on each value once in the fundamental region. It follows (Sec. 43, Theorem 14) that the most general simple automorphic function belonging to the group is a rational function of $\wp(z)^2$.

Differentiating (27) and (26), we have

$$\wp'(z_i) \frac{dz_i}{dz} = \frac{1}{K_\nu^2} \wp'(z), \quad \frac{dz_i}{dz} = K_\nu,$$

whence,

$$\Psi'(z_i) = \frac{1}{K_i^{2n}} \Psi'(z). \quad (28)$$

If $\nu = 3$ (Fig. 43), K_i is a cube root of unity and $K_i^{2n} = 1$. Then $\Psi'(z)$ is automorphic with respect to the group. $\Psi'(z)$ has a pole of the third order and takes on each value three times in the period parallelogram. The period parallelogram contains three copies of the fundamental region of the figure; hence, the function takes on each value once in the fundamental region. The most general simple automorphic function belonging to the group is a rational function of $\Psi'(z)$.

If $\nu = 6$ (Fig. 44), K_i is a sixth root of unity and $K_i^{2n} = \pm 1$. Then $\Psi'(z)^2$ is automorphic. This function has a pole of the sixth order and takes on each value six times in the period parallelogram. The period parallelogram contains six copies of the fundamental region of the figure; hence, the function takes on every value once in the fundamental region. The most general simple automorphic function belonging to the group is a rational function of $\Psi'(z)^2$.

III. THE GROUPS WITH TWO LIMIT POINTS

62. Determination of the Groups.—The remaining elementary groups—those with two limit points—are of less interest. We shall derive them briefly. Given such a group, so transformed that the limit points are zero and infinity. Then each transformation must carry a limit point either into itself or into the other limit point. The transformations (a) which have 0 and ∞ as fixed points; and (b) which carry 0 to ∞ and ∞ to 0, are readily found to have the following forms:

$$(a) z' = K_i z, \quad (b) z' = \frac{c_j}{z}. \quad (29)$$

The former are hyperbolic, elliptic, or loxodromic transformations, according to the value of K_i ; the latter are elliptic with the fixed points $\pm \sqrt{c_j}$.

Taking logarithms in (29) and setting $\log z = w$, $\log z' = w'$, $\log K_i = \Omega_j$, $\log c_j = \Omega'_j$, we have

(a) $w' = w + \Omega_j + 2\pi ni$, (b) $w' = -w + \Omega'_j + 2\pi ni$, (30)
the integers n and n' depending upon the values used for the logarithms. If the transformations in (29) form a group, so, also, do those in (30), and conversely. The group (29) is con-

tinuous or discontinuous according as (30) is continuous or discontinuous. Thus, (30) is continuous if it has arbitrarily small periods, $\Omega_j + 2\pi ni = \epsilon$. Then $\log z' = \log z + \epsilon$ and $z' = e^\epsilon z$, which differs arbitrarily little from the identity.

The groups (30) are of the type studied in Secs. 59 and 60 for $\nu = 2$. Suppose, first, that the transformations (b) do not appear. In (a) we can take one of the primitive periods to be a pure imaginary. Since $2\pi i$ is a period, this primitive period is of the form $\omega = 2\pi i/k$, where k is an integer. If there are no periods other than multiples of ω , the group (30)(a) is simply periodic and is generated by the transformation $w' = w + \omega$. The group (29)(a) is then generated by $z' = e^{\omega z} = e^{2\pi i/k} z$. It is a finite cyclic group and can be discarded, since it has no limit points.

If (30)(a) is a doubly periodic group, let ω, ω' be a pair of primitive periods, ω being the period found above. Then ω' is not a pure imaginary. The general transformation of the group is $w' = w + m\omega + n\omega'$ and the general transformation of the original group is $z' = e^{m\omega + n\omega'} z$. If $k = 1$, so that $\omega = 2\pi i$, then $e^{m\omega} = 1$ and the group is simply $z' = e^{n\omega'} z$. Putting $K = e^{\omega'}$, $K_1 = e^{\omega}$, we have the following types of groups:

$$(A) z' = K_i z; \quad (B) z' = K_i K_1^{n'} z \quad |K| \neq 1, K_1 = e^{2\pi i/k}$$

(A) is a hyperbolic or loxodromic cyclic group. (B) contains elliptic and loxodromic transformations and possibly also hyperbolic transformations.

If the transformations (b) appear in (30), they have the form (Equation (23) with $K = -1$)

$$w' = -w + b + m\omega + n\omega'.$$

The group can be transformed by a translation that carries b to the origin (which amounts in (30) to a transformation carrying $\sqrt{c}$ to 1); so we can take $b = 0$. From $\log z' = -\log z + m\omega + n\omega'$, we get $z' = e^{m\omega + n\omega'} / z$. We have then the following further groups:

$$(C) z' = K_i z, \quad z'' = \frac{K_i}{z}; \quad (D) z' = K_i K_1^{n'} z, \quad z'' = \frac{K_i K_1^{n'}}{z}.$$

For any values of K, K_1 such that $|K| \neq 1$ and $K_1 = e^{2\pi i/k}$, the sets of transformations (A)-(D) form groups of the kind sought.

The most general groups with two limit points are the groups (A)-(D) and their transforms by means of linear transformations.

CHAPTER VII

THE ELLIPTIC MODULAR FUNCTIONS

63. Certain Results from the Theory of Elliptic Functions.—The earliest automorphic functions to be studied were the elementary ones—the circular functions, the elliptic functions, and the rational or polyhedral automorphic functions. Of the non-elementary functions the so-called “elliptic modular functions” were extensively studied before the erection of a general theory of Fuchsian functions. By an elliptic modular function, or more briefly, a modular function, is meant a simple automorphic function belonging to the modular group or to one of its subgroups.¹

In Sec. 37, we constructed the fundamental region for the modular group. By transforming the group so as to reduce infinity to an ordinary point, we could set up automorphic functions by means of the theta series of Poincaré (Chap. V). Instead of following this process, however, we prefer to follow the historical order of development and to derive the modular functions in such a way as to bring

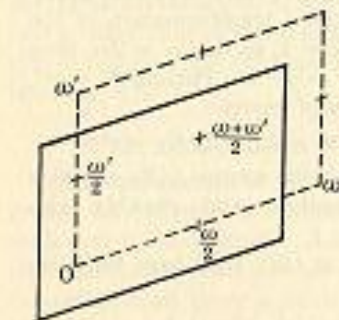


FIG. 45.

out their connections with the elliptic functions. To this end we recall certain of the properties of the Weierstrassian function $\wp(z)$.

Let ω, ω' be a pair of primitive periods so denominated that for the ratio

$$\tau = \frac{\omega'}{\omega} = z + iy, \quad (1)$$

¹ The most complete treatise on these functions is FRICKE-KLEIN, “Vorlesungen über die Theorie der elliptischen Modulfunktionen,” two volumes of some 1450 pages. See, also, VIVANTI, “Fonctions polyédriques et modulaires.” Brief accounts will be found in HURWITZ-COURANT, “Funktionentheorie,” 2nd ed., pp. 220–230, and in BIEBERBACH, “Lehrbuch der Funktionentheorie,” vol. 2, pp. 95–114.

y is positive; that is, which shall be called ω and which ω' is governed by the requirement that the angle $\omega\omega'$ shall be positive and less than π (Fig. 45). We use the symbol $\Omega = m\omega + m'\omega'$ for the general period. The elliptic function

$$\wp(z) = \frac{1}{z^2} + \sum' \left[\frac{1}{(z - \Omega)^2} - \frac{1}{\Omega^2} \right], \quad (2)$$

where the summation extends to all non-zero periods, and its derivative

$$\wp'(z) = -\frac{2}{z^3} - 2 \sum' \frac{1}{(z - \Omega)^3} \quad (3)$$

are connected by an algebraic relation (Theorem 12, Sec. 43). This relation is the following

$$\wp'(z)^2 = 4\wp(z)^3 - g_2\wp(z) - g_3, \quad (4)$$

where

$$g_2 = 60 \sum' \frac{1}{\Omega^4}, \quad g_3 = 140 \sum' \frac{1}{\Omega^6} \quad (5)$$

This relation is got most easily by combining $\wp(z)$ and $\wp'(z)$ into a polynomial in such a way that the pole at the origin disappears. Then the polynomial reduces to a constant (Theorem 10, Sec. 42). In the neighborhood of the origin, we have

$$\frac{1}{(z - \Omega)^2} = \frac{1}{\Omega^2} + \frac{2z}{\Omega^3} + \frac{3z^2}{\Omega^4} + \frac{4z^3}{\Omega^5} + \frac{5z^4}{\Omega^6} + \dots$$

We now insert this value in (2) and sum. We note that $\sum' \frac{1}{\Omega^2} = \sum' \frac{1}{\Omega^4} = \dots = 0$ by virtue of the fact that the term containing $\Omega = m\omega + m'\omega'$ is cancelled by the term containing $\Omega = -m\omega - m'\omega'$. We have

$$\wp(z) = \frac{1}{z^2} + 3z^2 \sum' \frac{1}{\Omega^4} + 5z^4 \sum' \frac{1}{\Omega^6} + \dots$$

or

$$\wp(z) = \frac{1}{z^2} + \frac{g_2}{20} z^2 + \frac{g_3}{28} z^4 + \dots$$

From this series, we have, at once, the following:

$$\wp'(z) = -\frac{2}{z^3} + \frac{g_2}{10} z + \frac{g_3}{7} z^3 + \dots,$$

$$\wp(z)^3 = \frac{1}{z^6} + \frac{3g_2}{20z^2} + \frac{3g_3}{28} + \dots,$$

$$\wp'(z)^2 = \frac{4}{z^6} - \frac{2g_2}{5z^2} - \frac{4g_3}{7} + \dots$$

On combining these equations we find

$$\wp'(z)^2 - 4\wp(z)^3 + g_2\wp(z) = -g_3 + cz^2 + \dots$$

The first member of this equation is an elliptic function with no singularity in the fundamental region; hence, it is a constant.

Setting $z = 0$, we see that its value is $-g_3$ and (4) is established.

The function $\wp'(z)$ satisfies the following:

$$\wp'\left(-\frac{\Omega}{2}\right) = \wp'\left(\frac{\Omega}{2}\right), \quad \wp'\left(-\frac{\Omega}{2}\right) = -\wp'\left(\frac{\Omega}{2}\right),$$

the first of which follows from the fact that $-\Omega/2$ and $\Omega/2$ differ by a period and the second from the fact that $\wp'(z)$ is an odd function of z . It follows that if $\wp'(z)$ is analytic at $\Omega/2$; that is, if $\Omega/2$ is a half period, then $\wp'(\Omega/2) = 0$. Three of these half periods lie in the period parallelogram. The three quantities

$$e_1 = \wp\left(\frac{\omega}{2}\right), \quad e_2 = \wp\left(\frac{\omega + \omega'}{2}\right), \quad e_3 = \wp\left(\frac{\omega'}{2}\right), \quad (6)$$

are then the roots of the equation

$$4t^3 - g_2t - g_3 = 0. \quad (7)$$

The three quantities e_1, e_2, e_3 are unequal. $\wp(z)$ takes on each value twice in the period parallelogram. Since $\wp'(z)$ vanishes at $\omega/2$, $\wp(z)$ takes on the value e_1 twice there, and so cannot take on the value e_1 elsewhere in the parallelogram. By similar reasoning e_2 and e_3 are unequal. The condition for equal roots in (7) we find readily to be the vanishing of the expression

$$\Delta = g_2^3 - 27g_3^2. \quad (8)$$

Hence,

$$\Delta \neq 0.$$

64. Change of the Primitive Periods.—We consider the problem of finding all pairs of primitive periods; that is, periods in terms of which we can express all other periods by adding integral multiples of the two periods. The periods

$$\omega_1' = a\omega' + b\omega, \quad \omega_1 = c\omega' + d\omega \quad (9)$$

a, b, c, d being integers, are primitive periods if, and only if, ω, ω' can be expressed as sums of integral multiples of ω_1, ω_1' . This condition is necessary, by definition. It is also sufficient. For, any period Ω , being expressible in integral multiples of ω, ω' , is then expressible in integral multiples of ω_1, ω_1' . Solving (9) we have

$$\omega' = c\omega_1' + \beta\omega_1, \quad \omega = \gamma\omega_1' + \delta\omega_1,$$

where, putting $D = ad - bc$,

$$\alpha = \frac{d}{D}, \quad \beta = -\frac{b}{D}, \quad \gamma = -\frac{c}{D}, \quad \delta = \frac{a}{D}.$$

Then

$$\alpha\delta - \beta\gamma = \frac{ad - bc}{D^2} = \frac{1}{D}.$$

If $\alpha, \beta, \gamma, \delta$ are integers, then $1/D$ is an integer; whence, $D = \pm 1$. Obviously, if $D = \pm 1$, the four quantities are integers. Hence, the necessary and sufficient conditions that (9) be primitive periods is that $ad - bc = \pm 1$.

The ratio of the new periods satisfies the relation

$$\tau_1 = \frac{\omega_1'}{\omega_1} = \frac{a\tau + b}{c\tau + d}. \quad (10)$$

Let us require that the periods be so named that in the expression $\tau_1 = x_1 + iy_1$, y_1 shall be positive. Then τ_1 as well as τ lies in the upper half plane. Now, the transformation (10) is non-loxodromic if $D > 0$ and is loxodromic if $D < 0$. In the former case, the upper half plane is carried into itself; in the latter it is carried into the lower half plane. We have then

$$ad - bc = +1. \quad (11)$$

The set of transformations (10) is the modular group.

65. The Function $J(\tau)$.—Expressing the quantities (5) and (8) in terms of the ratio τ , we have

$$\left. \begin{aligned} g_2 &= g_2(\omega, \omega') = 60 \sum' \frac{1}{(m\omega + m'\omega')^4} \\ &= \frac{60}{\omega^4} \sum' \frac{1}{(m + m'\tau)^4} = \frac{1}{\omega^4} g_2(1, \tau), \\ g_3 &= g_3(\omega, \omega') = \frac{40}{\omega^6} \sum' \frac{1}{(m + m'\tau)^6} = \frac{1}{\omega^6} g_3(1, \tau), \\ \Delta &= \Delta(\omega, \omega') = g_2(\omega, \omega')^3 - 27g_3(\omega, \omega')^2 = \frac{1}{\omega^{12}} \Delta(1, \tau). \end{aligned} \right\} \quad (12)$$

It is a well-known result that the series (5) converge absolutely for any pair of primitive periods ω, ω' whose ratio is not real. Hence, the series for $g_2(1, \tau)$ and $g_3(1, \tau)$ converge absolutely for any value of τ which is not real. We show below that these series converge uniformly in any closed region not containing real points. From this, it follows that $g_2(1, \tau)$ and $g_3(1, \tau)$ are analytic functions of τ in the whole upper half τ -plane. So also, then, is $\Delta(1, \tau)$.

We may establish the uniform convergence as follows: Let S be a closed region in the τ -plane not containing real points. Let ϵ be the minimum

distance from the boundary of S to the real axis; and let $|\tau| < N$ in S . We now show that $\epsilon > 0$ can be chosen sufficiently small that

$$|m + m'\tau| > \epsilon|m + m'i|; \text{ or } |m + m'\tau|^2 - \epsilon^2|m + m'i|^2 > 0,$$

for all values of m, m' in the summation and for all τ in S . Writing $\tau = x + iy$, the first member of the last inequality may be written

$$\begin{aligned} |m + m'x + im'y|^2 - \epsilon^2|m + m'i|^2 \\ = m^2 + 2mm'x + m'^2x^2 + m'^2y^2 - \epsilon^2m^2 - \epsilon^2m'^2 \\ = m^2 \left[1 - \frac{1}{k^2} - \epsilon^2 \right] + \left[\frac{1}{k}m + km'x \right]^2 \\ + m'^2[y^2 - (k^2 - 1)x^2 - \epsilon^2]. \end{aligned}$$

Here, since x is bounded, $|x| < N$, and $|y| \geq \eta$, we can take $k(>1)$ near enough to 1 and $\epsilon(>0)$ small enough that all the terms in the last member are positive whatever m and m' are; which establishes the desired inequality. The terms of the series for $g_2(1, \tau)$ are less in absolute value than the corresponding terms of the series of positive constant terms

$$\sum' \frac{1}{|m + m'i|^4 \epsilon^4},$$

which is known to converge. It follows that the series for $g_2(1, \tau)$ converges uniformly in S . The uniform convergence of the series for $g_3(1, \tau)$ is similarly established.

We now combine the quantities in (12) to form a function from which the factor ω cancels and which is, therefore, a function of the ratio τ alone:

$$J(\tau) = \frac{g_2^3}{\Delta} = \frac{g_2(1, \tau)^3}{\Delta(1, \tau)}. \quad (13)$$

The numerator and denominator are analytic functions of τ , and the denominator does not vanish, in the upper half plane. Hence, $J(\tau)$ is analytic in the whole upper half τ -plane.

Let ω, ω' of given ratio $\omega'/\omega = \tau$ generate a group. Then, the periods $\omega_1' = a\omega' + b\omega$, $\omega_2' = c\omega' + d\omega$, where a, b, c, d are any integers such that $ad - bc = 1$, are primitive periods for the group. Since the new periods generate the original group, they lead to the same $\mathfrak{P}$ -function and to the same constants g_2, g_3, Δ . Hence, the function J is unaltered. Now, the ratio of the new periods $\tau_1 = \omega_1'/\omega_2'$ is connected with that of the old by Equation (10); hence, we have, for any transformation of the modular group,

$$J\left(\frac{a\tau + b}{c\tau + d}\right) = J(\tau). \quad (14)$$

66. Behavior of $J(\tau)$ at the Parabolic Points.— $J(\tau)$ is analytic at all finite points of the fundamental region R_0 of Fig. 46. We now consider its behavior at the upper end of the region. The sides which meet at infinity are congruent by the transformation $\tau' = \tau + 1$. In accordance with Sec. 41, we make the change of variable

$$t = e^{2\pi i \tau} \quad (15)$$

This transformation maps the part of R_0 above the line DEF , with the equation $y = k > 1$, on the interior of the circle K whose equation is $|t| = e^{-2\pi k}$, congruent points on the two sides going into coincident points on a radius (Fig. 47). The function J takes on the same values at a point of this radius other than the

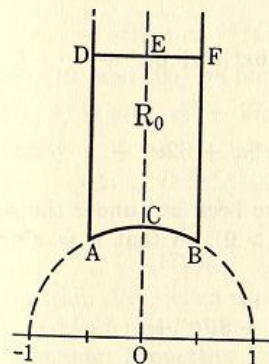


FIG. 46.

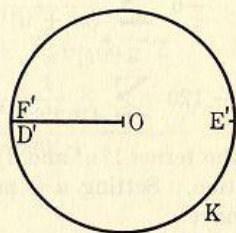


FIG. 47.

origin, when approached from the two sides, and so is single valued and analytic in K , with the possible exception of the origin. This is also true of $g_2(1, \tau)$ and $g_3(1, \tau)$, for each is unaltered by the transformation $\tau' = \tau + 1$. Thus,

$$\begin{aligned} g_2(1, \tau + 1) &= 60 \sum' \frac{1}{[m + m'(\tau + 1)]^4} \\ &= 60 \sum' \frac{1}{[m + m' + m'\tau]^4} \\ &= g_2(1, \tau), \end{aligned} \quad (16)$$

the terms of the second sum being merely a rearrangement of those of the sum defining $g_2(1, \tau)$. Similarly, $g_3(1, \tau + 1) = g_3(1, \tau)$.

In order to express $g_2(1, \tau)$ and $g_3(1, \tau)$ as functions of t , we shall employ two well-known expressions for the cotangent:¹

$$\pi \cot \pi u = \frac{1}{u} + \sum_{n=1}^{\infty} \left[\frac{1}{u+n} - \frac{1}{n} \right] \quad (17)$$

and

$$\cot \pi u = i \frac{w+1}{w-1}, \quad w = e^{2\pi i u},$$

whence,

$$\pi \cot \pi u = -\pi i [1 + 2w + 2w^2 + 2w^3 + \dots]. \quad (18)$$

We now equate the second members of (17) and (18) and differentiate, respectively, three and five times with respect to u , using each time the relation $dw/du = 2\pi i w$. We have the following results:

$$-6 \sum_{n=-\infty}^{\infty} \frac{1}{(m+n)^3} = -16\pi^4 [w + 8w^2 + \dots],$$

$$-120 \sum_{n=-\infty}^{\infty} \frac{1}{(m+n)^5} = 64\pi^4 [w + 32w^2 + \dots],$$

where the terms $1/n^3$ and $1/n^5$ have been put under the sign of summation. Setting $u = m'\tau$ ($m' > 0$), so that $w = e^{2\pi i m'\tau} = t^{m'}$ we have

$$\sum_{n=-\infty}^{\infty} \frac{1}{(m+m'\tau)^3} = \frac{8\pi^4}{3} [t^{m'} + 8t^{2m'} + \dots],$$

$$\sum_{n=-\infty}^{\infty} \frac{1}{(m+m'\tau)^5} = -\frac{8\pi^4}{15} [t^{m'} + 32t^{2m'} + \dots].$$

We can write²

$$\begin{aligned} g_2(1, \tau) &= 60 \left[\sum_{n=-\infty}^{\infty} \frac{1}{m^4} + 2 \sum_{n=1}^{\infty} \sum_{m=-\infty}^{\infty} \frac{1}{(m+m'\tau)^4} \right] \\ &= 60 \left[\frac{\pi^4}{45} + \frac{16\pi^4}{3} \sum_{m'=1}^{\infty} (t^{m'} + 8t^{2m'} + \dots) \right]. \end{aligned}$$

¹ For the first, see OSGOOD, W. F., "Lehrbuch der Funktionentheorie," 2nd ed., p. 507.

² The values of $\sum' \frac{1}{m^4}$ and $\sum' \frac{1}{m^6}$ which are used here and in (20) are readily got from (17) by expanding $\pi \cot \pi u = \frac{1}{u}$ in powers of u , differentiating three and five times, and setting $u = 0$.

Now, the functions in the last summation are analytic in the whole interior of K , including the origin. Since the series converges uniformly on the circumference of K , the sum is an analytic function throughout K . Also, we can collect the powers of t . We have

$$g_2(1, \tau) = \pi^4 \left[\frac{1}{45} + 320t + \dots \right]. \quad (19)$$

Similarly,

$$\begin{aligned} g_3(1, \tau) &= 140 \left[\sum_{n=-\infty}^{\infty} \frac{1}{m^6} + 2 \sum_{n=1}^{\infty} \sum_{m=-\infty}^{\infty} \frac{1}{(m+m'\tau)^6} \right] \\ &= 140 \left[\frac{2\pi^6}{945} - \frac{16\pi^6}{15} \sum_{m'=1}^{\infty} (t^{m'} + 32t^{2m'} + \dots) \right] \\ &= \pi^6 \left[\frac{2}{315} - 448t + \dots \right]. \end{aligned} \quad (20)$$

From (19) and (20) we find

$$\Delta(1, \tau) = g_2(1, \tau)^3 - 27g_3(1, \tau)^2 = \pi^{12} [4096t + \dots], \quad (21)$$

and finally

$$\begin{aligned} J(\tau) &= \frac{g_2(1, \tau)^3}{\Delta(1, \tau)} = \frac{\pi^{12} (\frac{1}{45} + 320t + \dots)^3}{\pi^{12} (4096t + \dots)} \\ &= \frac{1}{1728t} + c_0 + c_1t + \dots \end{aligned} \quad (22)$$

The function $J(\tau)$, when expressed in terms of t , thus has a pole of the first order at $t = 0$.

§7. Further Properties of $J(\tau)$.—The function $J(\tau)$ has a single pole of the first order in R_0 , according to the convention of Sec. 42. It, therefore, takes on each value once in R_0 (Theorem 11, Sec. 42).

THEOREM 1.—The function $J(\tau)$ takes on each value once, and only once, in the fundamental region.

Applying Theorem 14, Sec. 43, we determine all functions which are automorphic with respect to the modular group and have no other singularities than poles in the fundamental region.

THEOREM 2.—The most general simple automorphic function belonging to the modular group is a rational function of $J(\tau)$.

We next consider the values of $J(\tau)$ at the vertices A, B, C of R_0 . At A we have $\tau = \rho = -\frac{1}{2} + \frac{1}{2}i\sqrt{3}$. Here, ρ is a cube root of unity, so $\rho^3 = 1$ and $\rho^2 + \rho + 1 = 0$. If, in

$$g_2(1, \rho) = 60 \sum' \frac{1}{(m+m'\rho)^4}$$

we multiply numerator and denominator by ρ^2 , thus inserting a factor ρ^2 in each factor in the denominator, we have

$$\begin{aligned} g_2(1, \rho) &= 60\rho^2 \sum' \frac{1}{(m\rho^2 + m')^4} \\ &= 60\rho^2 \sum' \frac{1}{(m' - m - m\rho)^4} = \rho^2 g_2(1, \rho), \end{aligned}$$

for the terms of the last summation are merely those of the original sum arranged in a different order. From this, since $\rho^2 \neq 1$, we have

$$g_2(1, \rho) = 0, \quad J(\rho) = 0. \quad (23)$$

J vanishes also at the congruent point B .

At C we have $\tau = i$. Then,

$$\begin{aligned} g_2(1, i) &= 140 \sum' \frac{1}{(m + m'i)^6} \\ &= 140i^6 \sum' \frac{1}{(mi - m')^6} = -g_2(1, i). \end{aligned}$$

From this we have

$$g_2(1, i) = 0, \quad J(i) = 1. \quad (24)$$

We propose now to find all points of R_0 at which $J(\tau)$ is real. We consider the reflection of a point τ in the imaginary axis; namely $\tau' = -\bar{\tau}$, where, as hitherto, the bar indicates the conjugate imaginary. We have

$$g_2(1, \tau') = 60 \sum' \frac{1}{(m - m'\tau')^4}; \quad \bar{g}_2(1, \tau) = 60 \sum' \frac{1}{(m + m'\tau)^4};$$

so, $g_2(1, \tau') = \bar{g}_2(1, \tau)$. Similarly, $g_3(1, \tau') = \bar{g}_3(1, \tau)$; and

$$J(\tau') = \bar{J}(\tau). \quad (25)$$

There are two cases in which $J(\tau') = J(\tau)$. If τ lies on the imaginary axis, it coincides with its reflection, and $\tau' = \tau$. If τ lies on the boundary of R_0 , its reflection τ' lies at a congruent point on the boundary at which J has the same value as at τ . In both cases we have

$$J(\tau) = \bar{J}(\tau),$$

and $J(\tau)$ is real. That is, $J(\tau)$ is real on the boundary of R_0 and on the imaginary axis.

There are no further points of R_0 at which $J(\tau)$ is real. For, if the function is real at any other point τ_1 of R_0 , we have $J(\tau_1) = \bar{J}(\tau_1) = J(\tau'_1)$; and the function takes on the same value

at the two distinct inner points τ_1, τ'_1 of R_0 , which is contrary to Theorem 1.

It follows from the preceding that the imaginary part of the function $J(\tau) = u + iv$ has always the same sign in each of the halves into which the imaginary axis divides R_0 . If $\tau_1 > 0$ and $v_{\tau_1} < 0$ at two points of R_0 on the same side of the imaginary axis, and if τ trace a curve λ from τ_1 to τ_2 , lying in R_0 and not meeting the imaginary axis, then $v = 0$ at some point of λ . At this point, the function is real, which is contrary to what we have just proved.

It remains to find in which half v is greater than zero. As τ moves from A to C along the boundary of R_0 , $J(\tau)$ moves from 0 to 1 to the right along the real axis in the J plane. Points in the neighborhood of the first path and to the left of it go into points in the neighborhood of the second path and to the left of it; that is, into points of the upper half J -plane. Then $v > 0$ in the left half of R_0 .

We collect our results into the following theorem:

THEOREM 3.—Writing $J(\tau)$ in terms of its real and imaginary parts, $J(\tau) = u + iv$; then $v > 0$ at the inner points of R_0 to the left of the imaginary axis; $v < 0$ at the inner points of R_0 to the right of the imaginary axis; $v = 0$ on the boundary of R_0 and on the imaginary axis. Also, $v_{\tau'} = -v_{\tau}$, where τ' is the reflection of τ in the imaginary axis.

The last statement of the theorem is an immediate consequence of (25).

The function $z = J(\tau)$ maps the left half of R_0 in a one-to-one manner on the upper half z -plane and maps the right half similarly on the lower half-plane. As τ moves upward along the imaginary axis from C to ∞ , z moves from 1 to the right to $+\infty$. The two halves of the z -plane are joined to the right of $z = 1$. R_0 is mapped on the z -plane bounded by a slit which extends along the real axis from $z = 1$ to $z = -\infty$.

68. The Function $\lambda(\tau)$.—Of the elliptic modular functions belonging to a subgroup of the modular group, and not automorphic with respect to the whole group, we shall treat here only one. The function which we shall consider is defined as follows,

$$\lambda(\tau) = \frac{e_2 - e_3}{e_1 - e_3}, \quad (26)$$

where e_1, e_2, e_3 have the definitions given in Equation (6).

We note first that λ is, in fact, a function of the ratio τ alone. We write $\mathfrak{P}(z; \omega, \omega')$ to call attention to the dependence of the $\mathfrak{P}$ -function on the periods. Since all the terms in the series (2) defining $\mathfrak{P}$ are of the (-2) nd degree in z, ω, ω' , we have

$$\mathfrak{P}(z; \omega, \omega') = k^2 \mathfrak{P}(kz; k\omega, k\omega'), \quad k \neq 0. \quad (27)$$

Applying this to Equations (6), taking $k = 1/\omega$, we have

$$\left. \begin{aligned} e_1 &= \mathfrak{P}\left(\frac{\omega}{2}; \omega, \omega'\right) = \frac{1}{\omega^2} \mathfrak{P}\left(\frac{1}{2}; 1, \tau\right), \\ e_2 &= \mathfrak{P}\left(\frac{\omega + \omega'}{2}; \omega, \omega'\right) = \frac{1}{\omega^2} \mathfrak{P}\left(\frac{1 + \tau}{2}; 1, \tau\right), \\ e_3 &= \mathfrak{P}\left(\frac{\omega'}{2}; \omega, \omega'\right) = \frac{1}{\omega^2} \mathfrak{P}\left(\frac{\tau}{2}; 1, \tau\right). \end{aligned} \right\} \quad (28)$$

On forming the quotient in (26), ω cancels, and λ is a function of τ alone.

The uniform convergence of the series that define the functions of τ appearing in (28) can be proved for any closed region in the τ -plane which does not contain points of the real axis. This proof follows the lines of the convergence proof in Sec. 65 and will not be given here. These functions are then analytic in the upper half plane. Since e_1, e_2, e_3 are unequal, $\lambda(\tau)$ is an analytic function of τ , and nowhere takes on the value 0 or the value 1, in the upper half τ -plane.

Consider the group generated by a pair of periods ω, ω' whose ratio ω'/ω is τ . Let ω_1, ω'_1 , Equation (9), be a second pair of primitive periods with the condition that $ad - bc = +1$. The ratio of the new periods is given by (10). For the new periods, we have

$$\begin{aligned} e_1' &= \mathfrak{P}\left(\frac{c\omega' + d\omega}{2}\right), \quad e_2' = \mathfrak{P}\left(\frac{c\omega' + d\omega + a\omega' + b\omega}{2}\right), \\ e_3' &= \mathfrak{P}\left(\frac{a\omega' + b\omega}{2}\right). \end{aligned} \quad (29)$$

Since we are dealing with the same group and, hence, have the same $\mathfrak{P}$ -function, we have the same constants as before. They are possibly arranged in a different order. And a rearrangement of the constants in (26) alters the value of λ , in general.

If now $(c\omega' + d\omega)/2$ and $\omega/2$ differ by a period, that is, if $(c\omega' + (d-1)\omega)/2$ is a period, we have

$$\mathfrak{P}\left(\frac{c\omega' + d\omega}{2}\right) = \mathfrak{P}\left(\frac{\omega}{2}\right), \text{ or } e_1' = e_1.$$

This occurs if, and only if, c is even and d is odd. Similarly, if $(a\omega' + b\omega)/2$ and $\omega'/2$ differ by a period, that is, if $((a-1)\omega' + b\omega)/2$ is a period, we have $e_3' = e_3$. This occurs if a is odd and b is even. If the preceding conditions hold, the remaining roots are equal, $e_2' = e_2$, and λ is unchanged.

Now if b and c are even, $b = 2b', c = 2c'$, then, a and d are necessarily odd, as a consequence of the relation $ad - bc = 1$; and we have

$$\lambda\left(\frac{a\tau + 2b'}{2c'\tau + d}\right) = \lambda(\tau), \quad ad - 4b'c' = 1. \quad (30)$$

$\lambda(\tau)$ is thus unaltered by all the transformations of the subgroup treated in Sec. 38 and whose fundamental region was found in Fig. 30. (For convenience this region is repeated in Fig. 48.)

69. The Relation between $\lambda(\tau)$ and $J(\tau)$.—If we make a transformation of the modular group which does not belong to the subgroup of (30), there is an interchange of the roots e_1, e_2, e_3 . All values of (26) which can result are contained in the following table:

$$\left. \begin{aligned} \frac{e_2 - e_3}{e_1 - e_3} &= \lambda, & \frac{e_2 - e_1}{e_3 - e_1} &= \frac{1}{1 - \lambda}, & \frac{e_1 - e_2}{e_3 - e_2} &= \frac{\lambda - 1}{\lambda}, \\ \frac{e_1 - e_3}{e_2 - e_3} &= \frac{1}{\lambda}, & \frac{e_3 - e_2}{e_1 - e_2} &= \frac{\lambda}{\lambda - 1}, & \frac{e_2 - e_1}{e_3 - e_1} &= 1 - \lambda. \end{aligned} \right\} \quad (31)$$

These six transformations of λ constitute a group, namely, the group of the anharmonic ratios (Sec. 36).

If, now, we form a rational symmetric function of the six quantities (31), we have a simple automorphic function $F(\tau)$ belonging to the modular group. For, in the first place, any modular transformation merely interchanges the quantities e_1, e_2, e_3 in a certain way and does not alter a symmetric function of these quantities. And, in the second place, such a symmetric function of the roots can be expressed as a rational function of the coefficients $g_2(\omega, \omega')$, $g_3(\omega, \omega')$, and, hence, also of $g_2(1, \tau)$, $g_3(1, \tau)$. The latter two functions have no singularities other than poles in R_2 (including the parabolic point); whence, a rational function of them has no singularities other than poles.

Then $F(\tau)$ is a simple automorphic function. It follows from Theorem 2 that $F(\tau)$ is a rational function of $J(\tau)$.

We shall consider the following symmetric function:

$$F(\tau) = (\lambda + 1) \left(\frac{1}{1-\lambda} + 1 \right) \left(\frac{\lambda-1}{\lambda} + 1 \right) \left(\frac{1}{\lambda} + 1 \right) \left(\frac{\lambda}{\lambda-1} + 1 \right) \\ = - \frac{(\lambda+1)^2(2-\lambda)^2(2\lambda-1)^2}{\lambda^2(1-\lambda)^2}. \quad (32)$$

Putting this in terms of e_1, e_2, e_3 , we have

$$F(\tau) = - \frac{(e_2 + e_1 - 2e_3)^2(e_2 + e_3 - 2e_1)^2(e_1 + e_3 - 2e_2)^2}{(e_1 - e_2)^2(e_2 - e_3)^2(e_3 - e_1)^2}. \quad (33)$$

From Equation (7) we have $e_1 + e_2 + e_3 = 0$, $e_1 e_2 e_3 = g_3/4$. Using these relations, we can write the numerator of (33) in the form

$$(-3e_2)^2(-3e_1)^2(-3e_3)^2 = \frac{3^6}{4^3} g_3^2.$$

The denominator of (33) is equal to the discriminant of (7), save for a constant factor. We have, in fact,

$$(e_1 - e_2)^2(e_2 - e_3)^2(e_3 - e_1)^2 = \frac{1}{16}(g_2^3 - 27g_3^2).$$

Then,

$$F(\tau) = - \frac{3^6 g_3^2}{g_2^3 - 27g_3^2} = 27[1 - J(\tau)]. \quad (34)$$

From (34) and (32) we have, with a little calculation,

$$J(\tau) = 1 - \frac{F(\tau)}{27} = \frac{4}{27} \frac{(1-\lambda+\lambda^2)^2}{\lambda^2(1-\lambda)^2}. \quad (35)$$

70. Further Properties of $\lambda(\tau)$.—We first note that for each point of B_2 there exist five other points which are congruent by the modular group. In Fig. 49 we have superposed the modular division of the plane on B_2 . Representing the left half of B_0 and the triangles congruent to it by the + sign and the right half and its congruent triangles by the - sign, we observe that there are exactly six copies of each. Then $J(\tau)$ takes on each value six times in B_0 .

We next show that $\lambda(\tau)$ is a simple automorphic function. The only doubtful matter is its behavior at the parabolic points. Consider the parabolic point at infinity. If, in accordance with Sec. 41, we make the change of variable $t_1 = e^{\tau/\omega}$ (the sides of B_0 being congruent by the transformation $\tau' = \tau + 2$), the upper

end of B_1 is mapped on a region similar to Fig. 47, and λ goes over into a function of t_1 which is analytic in the neighborhood of the origin. If λ has an essential singularity at $t_1 = 0$, it takes on certain values an infinite number of times. Then, from (35), J takes on certain values an infinite number of times in B_0 , which is impossible. Hence, λ is analytic at $t_1 = 0$ or has a pole there. Similar remarks apply to the other parabolic points. Then $\lambda(\tau)$ is a simple automorphic function. It, therefore, takes on each value the same number of times.

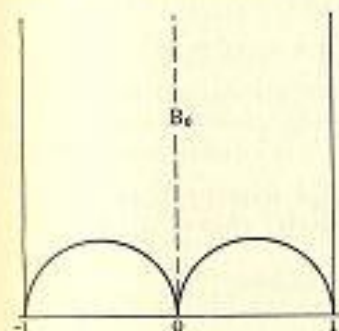


FIG. 48.

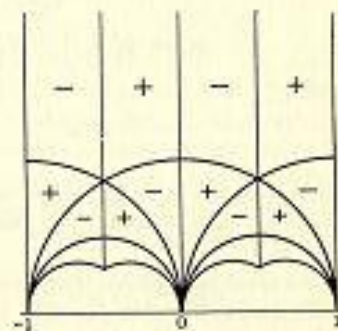


FIG. 49.

The exceptional values 0, 1, ∞ , which are taken on nowhere in the upper half plane, are necessarily taken on in the parabolic points 0, ± 1 , ∞ . The points $+1$ and -1 are congruent and bear the same value. We shall now determine in which points these values are taken on. One method of procedure would be to seek the limit approached by $\lambda(\tau)$ as τ approaches the parabolic point from within the region. An easier method is the following:

Let us make the change of periods (see (9) and (10))

$$\omega_1' = \omega' + \omega, \quad \omega_1 = \omega, \quad \tau_1 = \tau + 1.$$

The new constants are

$$e_1' = \wp\left(\frac{\omega_1}{2}\right) = \wp\left(\frac{\omega}{2}\right) = e_1,$$

$$e_2' = \wp\left(\frac{\omega_1'}{2}\right) = \wp\left(\frac{\omega + \omega'}{2}\right) = e_2, \quad e_2' = e_2,$$

and

$$\lambda' = \frac{e_2' - e_1'}{e_1' - e_1} = \frac{e_2 - e_1}{e_1 - e_1} = \frac{\lambda}{\lambda - 1}.$$

Now, when we make this change of primitive periods, $\tau = 0, -1, \infty$ go into $\tau' = 1, 0, \infty$, respectively; and $\lambda = 0, 1, \infty$ go into $\lambda' = 0, \infty, 1$, respectively. Since $\tau = \infty$ is unaltered, the corresponding value of λ is unaltered. Hence, $\lambda(\infty) = 0$.

Again, make the change of periods

$$\omega_1' = -\omega, \omega_2' = \omega, \tau_1' = -\frac{1}{\tau}.$$

Then,

$$e_1' = \wp\left(\frac{\omega_1'}{2}\right) = \wp\left(\frac{\omega}{2}\right) = e_1,$$

$$e_2' = \wp\left(\frac{\omega_2'}{2}\right) = \wp\left(-\frac{\omega}{2}\right) = e_1, \quad e_3' = e_2,$$

and

$$\lambda' = \frac{e_2 - e_1}{e_3 - e_1} = 1 - \lambda.$$

Here, $\tau = 0, -1, \infty$ go into $\tau_1' = \infty, 1, 0$, respectively; and $\lambda = 0, 1, \infty$ go into $\lambda' = 1, 0, \infty$, respectively; whence, $\lambda(\pm 1) = \infty$, $\lambda(0) = 1$.

We shall now prove the following theorem:

THEOREM 4.—The function $\lambda(\tau)$ takes on each value once, and only once, in the fundamental region B_1 .

Each simple automorphic function belonging to the subgroup is a rational function of $\lambda(\tau)$.

We find, first, the order of the zero of $\lambda(\tau)$ at the parabolic point at ∞ . As $\tau \rightarrow \infty$ in B_0 , $\lambda \rightarrow 0$, and, from (35),

$$\lambda^2 J \rightarrow \frac{1}{2} \tau^2. \quad (36)$$

Changing the variable to $t_1 = e^{\pi i \tau}$, we have from (22) and (15)

$$J = \frac{1}{1728t_1^2} + c_0 + c_1 t_1 + \dots = \frac{1}{1728t_1^2} + c_0 + c_1 t_1^2 + \dots$$

J thus has a pole of the second order at $t_1 = 0$. Hence, from (36), λ has a zero of the first order. Since $\lambda(\tau)$ has one, and only one, zero in B_0 , it takes on every value exactly once.

The second statement of the theorem is a consequence of Theorem 14, Sec. 43.

The reality of $\lambda(\tau)$ can be treated by the method used in studying the reality of $J(\tau)$. We find here, also, that $\lambda(-\bar{\tau}) = \bar{\lambda}(\tau)$; and by repeating practically verbatim the reasoning in connection with Theorem 3, we show that $\lambda(\tau)$ is real on the

boundary of B_0 and on the imaginary axis. The right half of B_0 corresponds to the upper half λ -plane.

We close the chapter with a general theorem concerning modular functions.

THEOREM 5.—Given a subgroup of the modular group whose fundamental region consists of k copies of the fundamental region R_0 of the modular group. Then any simple automorphic function $f(\tau)$ belonging to the subgroup is connected with $J(\tau)$ by a relation of the form

$$\Phi(f, J) = 0,$$

where Φ is a polynomial of degree not exceeding k in f .

The fact that an algebraic relation between f and J exists follows from Theorem 12, Sec. 43, since $J(\tau)$ is, obviously, a simple automorphic function with respect to the subgroup. To each value J_0 of J , there correspond not more than k distinct values of f ; namely, the values of $f(\tau)$ at the k points of the fundamental region at which $J(\tau) = J_0$. That is, the irreducible relation connecting f and J is of degree k at most in f .

The equation (35) connecting J and λ is an example.